

Europäisches Patentamt

European Patent Office

Office européen des brevets



(11) **EP 0 823 683 A1**

(12)

EUROPEAN PATENT APPLICATION

published in accordance with Art. 158(3) EPC

(43) Date of publication:
11.02.1998 Bulletin 1998/07

(51) Int. Cl.⁶: **G06F 3/033**

(21) Application number: 96912232.4

(86) International application number:
PCT/JP96/01124

(22) Date of filing: 25.04.1996

(87) International publication number:
WO 96/34332 (31.10.1996 Gazette 1996/48)

(84) Designated Contracting States:
DE FR GB

(30) Priority: 28.04.1995 JP 105551/95
19.10.1995 JP 270828/95
01.11.1995 JP 284768/95

(71) Applicant:
MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD.
Kadoma-shi, Osaka 571 (JP)

(72) Inventors:
• MARUNO, Susumu
Osaka 534 (JP)

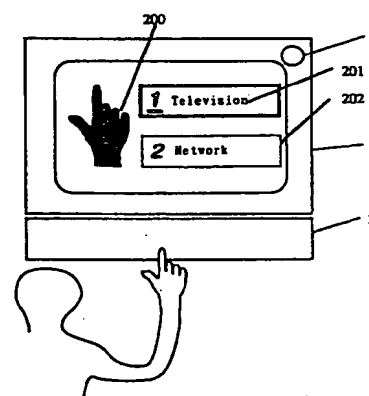
• IMAGAWA, Taro
Osaka 573 (JP)
• MORIYA, Michiyo
Osaka 570 (JP)
• KONDO, Kenji
Syoueiryoku 428
Osaka 572 (JP)

(74) Representative:
Crawford, Andrew Birkby et al
A.A. THORNTON & CO.
Northumberland House
303-306 High Holborn
London WC1V 7LE (GB)

(54) INTERFACE DEVICE

(57) The invention relates to an interface apparatus for making input and output of appliances having display such as computer, word processor, information appliance and television, comprising recognizing means for recognizing the shape or move of the hand of an operator, display means for displaying the features of the shape or move of the hand recognized by the recognizing means as special shape in the screen, and control means for controlling the information displayed in the screen by the special shape displayed in the screen by the display means, wherein the two-dimensional or three-dimensional information displayed in the screen can be selected, indicated or moved only by changing the shape or moving the hand, so that the interface apparatus of very excellent controllability and high diversity may be presented.

Fig. 1



EP 0 823 683 A1

Description

BACKGROUND OF THE INVENTION

The present invention relates to an interface apparatus for input and output of information apparatus such as computer and word processor and appliance having a display such as television.

In a kind of conventional interface apparatus, it is designed to display a cursor at a coordinate position detected by the mouse on a display screen, for adding some other information to the information in the display device, or changing or selecting the displayed information.

Fig. 30 shows an outline of this conventional interface apparatus. In Fig. 30, reference numeral 501 denotes a host computer, and 502 is a display, and virtual operation buttons 503, 504, 505 are displayed in the display 502 by the host computer 501. Reference numeral 506 represents a mouse cursor, and the host computer 501 controls the display so as to move in the screen in synchronism with the move of the mouse 507, on the basis of the moving distance of the mouse 507 detected by the mouse 507. As the user moves the mouse 507, the mouse cursor 506 is moved to the position of a desired virtual operation button in the display screen, and by pressing a switch 508 on the mouse 507, an operation button is selected so as to instruct action to the host computer 501.

In this conventional construction, however, the mouse or the input device is necessary in addition to the main body of the appliance, and a table or area for manipulating the mouse is also needed, which is not suited to portable information appliance or the like. Besides, by manipulation through the mouse, it is not a direct and intuitive interface.

SUMMARY OF THE INVENTION

It is an object of the invention to present an interface apparatus capable of manipulating an appliance easily without requiring input device such as keyboard and mouse. It is other object thereof to present an interface apparatus further advanced in the ease of manipulation of indicating or catching the display object by judging interactions along the intent of the operator sequentially and automatically.

In structure, the invention provides an interface apparatus comprising recognizing means for recognizing the shape of a hand of an operator, display means for displaying the features of the shape of the hand recognized by the recognizing means on the screen as a special shape, and control means for controlling the information displayed in the screen by the special shape displayed in the screen by the display means, whereby the information displayed in the screen can be controlled only by varying the shape of the hand.

It is a further object to present an interface appara-

tus much superior in ease of manipulation by recognizing also the move of the hand. To recognize the move, a frame memory for saving the image picking up the shape or move of the hand, and a reference image memory for storing the image taken before the image saved in the frame memory as reference image are provided, and it is achieved by depicting the difference between the image in the frame memory and the reference image stored in the reference image memory. In other method of recognition, the shape or move of the hand of the user in the taken image is depicted as the contour of the user, and its contour is traced, and the relation between the angle of the contour line and the length of contour line, that is, the contour waveform is calculated and filtered, and the shape waveform expressing the specified shape is generated.

Moreover, comprising cursor display means for displaying a feature of the shape of a hand on the screen as a special shape and manipulating as cursor, means for storing the relation with display object other than cursor displays the coordinates and shape of the representative point representing the position of the display object other than cursor display, and means for calculating and judging the interaction of the cursor display and the display object, manipulation is realized smoothly by the interactions along the intent of the operator when gripping the displayed virtual object in the case of display of cursor display as virtual manipulator.

In the interface apparatus thus constructed, as the user faces the recognizing means and shows, for example, a hand, the special shape corresponding to the shape of the hand is displayed as an icon in the screen for screen manipulation, so that control according to the icon display is enabled.

Or when instructed by hand gesture, the given hand gesture is displayed as a special shape set corresponding to the shape of the hand on the display screen, and its move is also displayed, and, for example, a virtual switch or the like displayed on the display screen can be selected by the hand gesture, or the display object displayed on the screen can be grabbed or carried depending on the purpose, and therefore without requiring mouse or other input device, a very simple manipulation of appliance is realized.

It is further possible to realize the interface much enhanced in the ease of manipulation by sequentially and automatically judging the interaction with the display object desired to be operated by the virtual manipulator according to the intent of operation of the operator, as the special shape set corresponding to the shape of the hand works as virtual manipulator aside from the mere cursor.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is an appearance drawing of an interface apparatus in a first embodiment of the invention;
Fig. 2 is a detailed block diagram of the interface

apparatus in the same embodiment of the invention;

Fig. 3 is a diagram showing an example of shape of hand judged by the interface apparatus in the same embodiment of the invention;

Fig. 4 is a diagram showing an example of shape identifying means of the interface apparatus in the same embodiment of the invention;

Fig. 5 is a diagram showing an example of operation by an image difference operation unit in the same embodiment;

Fig. 6 is a diagram showing an example of icon generated by an icon generating unit in the same embodiment;

Fig. 7 is an appearance drawing showing an operation example of the interface apparatus of the same embodiment;

Fig. 8 is an appearance drawing of an interface apparatus in a second embodiment of the invention; Fig. 9 is detailed block diagram of the interface apparatus in the second embodiment of the invention;

Fig. 10 is a diagram showing an example of shape of hand judged by the interface apparatus of the same embodiment;

Fig. 11 is a diagram showing an example of motion recognizing unit of the interface apparatus of the same embodiment;

Fig. 12 is a diagram showing an example of operation by an image difference operation unit in the same embodiment;

Fig. 13 is a diagram showing an operation example of the same embodiment;

Fig. 14 is a detailed block diagram of an interface apparatus in a third embodiment of the invention;

Fig. 15 is a diagram showing an example of motion recognizing unit of the interface apparatus in the third embodiment of the invention;

Fig. 16 (A) to (D) are diagrams showing examples of icon displayed on a display screen by the interface apparatus of the same embodiment;

Fig. 17 is a diagram showing operation of motion recognizing unit of the interface apparatus in the same embodiment of the invention;

Fig. 18 is a diagram showing operation of motion recognizing unit of the interface apparatus in the same embodiment of the invention;

Fig. 19 is a diagram showing operation of motion recognizing unit of the interface apparatus in the same embodiment of the invention;

Fig. 20 is a diagram showing operation of motion recognizing unit of the interface apparatus in the same embodiment of the invention;

Fig. 21 is a diagram showing an interface apparatus explaining a fourth embodiment;

Fig. 22 (a) is a diagram showing an open state of cursor in an example of a cursor used in the interface apparatus of the same embodiment;

(b) is a diagram showing a closed state of the same embodiment;

(c) is a diagram showing an open state of cursor in an example of a cursor used in the interface apparatus of the same embodiment;

(d) is a diagram showing a closed state of the same embodiment;

(e) is a diagram showing an open state of cursor in an example of a cursor used in the interface apparatus of the same embodiment;

(f) is a diagram showing a closed state of the same embodiment;

Fig. 23 (a) is a diagram showing the shape of an example of a virtual object used in the interface apparatus of the same embodiment;

(b) is a diagram showing the shape of other example of a virtual object used in the interface apparatus of the same embodiment;

Fig. 24 (a) is a front view showing configuration of cursor and virtual object in a virtual space;

(b) is a side view showing configuration of cursor and virtual object in a virtual space;

Fig. 25 is a diagram showing a display example of virtual space for explaining the embodiment;

Fig. 26 is a block diagram showing an example of the interface apparatus of the same embodiment;

Fig. 27 (a) is a diagram showing an example of input device in input means used in the interface apparatus of the same embodiment;

(b) is a diagram showing an example of input device in input means used in the interface apparatus of the same embodiment;

(c) is a diagram showing an example of input device in input means used in the interface apparatus of the same embodiment;

Fig. 28 (a) is a diagram showing an example of image of a hand taken by a camera in the same embodiment;

(b) is a diagram showing a binary example of image of a hand taken by a camera in the same embodiment;

Fig. 29 (a) is a diagram showing an example of image displayed by display means used in the interface apparatus in the same embodiment of the invention;

(b) is a diagram showing a second example of the display screen;

(c) is a diagram showing a third example of the display screen;

(d) is a diagram showing a fourth example of the display screen; and

Fig. 30 is an explanatory diagram for explaining a conventional interface apparatus.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

(First embodiment)

A first embodiment of the invention relates to an interface comprising recognizing means such as image pickup device for recognizing the shape of a hand of the operator, display means for displaying the feature of the shape of the hand recognized by the recognizing means on a screen as a special shape by an icon or the like, and control means for controlling the information displayed on the screen by varying the shape of the hand by operating the special shape such as icon displayed on the screen by the display means as the so-called cursor.

Fig. 1 shows the appearance of the first embodiment of the interface apparatus of the invention. Reference numeral 1 denotes a host computer, 2 is a display unit, and 3 is a CCD camera for picking up an image. The CCD camera 3 has the pickup surface located in the same direction as the display direction, so that the shape of the hand of the user can be picked up when the user confronts the display screen. On the display, menu 201, 202, and icon 200 reflecting the shape of the hand are displayed.

Fig. 2 is a detailed block diagram of the invention. The image fed in from the CCD camera is stored in a frame memory 21. In a reference image memory 25, a background image not including person taken previously is stored as reference image. The reference image may be updated whenever as required.

Shape identifying means 22 depicts the difference of the image saved in the frame memory and the image stored in the reference image memory, and removes the background image from the image, depicts, for example, the portion corresponding to the hand of the user, and judges if the shape is, for example, one finger as shown in Fig. 3 (A), two fingers as shown in Fig. 3 (B), or three fingers as shown in Fig. 3 (C).

Fig. 4 shows a detailed example of the shape identifying means 22, which comprises an image difference operation unit 221, a contour depicting unit 222, and a shape identifying unit 223.

The image difference operation unit 221 calculates the difference of the image saved in the frame memory and the image stored in the reference image memory as mentioned above. As a result, the object to be detected, for example, the user, can be separated from the background portion. For example, when the image difference operation unit 221 is composed of a simple subtraction circuit, as shown in Fig. 5, only the portion of the hand of the user in the image in the frame memory can be depicted. The contour depicting unit 222 depicts the contour shape of the object existing in the image as a result of operation by the image difference operation unit 221. As a practical method, for example, by depicting the edge of the image, the contour shape may be

easily depicted.

The shape identifying unit 223 identifies specifically the contour shape of the hand depicted by the contour depicting unit 222, and judges if the shape is, for example, one finger as shown in Fig. 3 (A) or two fingers as shown in Fig. 3 (B). As the shape identifying method, for example, template matching, matching technique with shape model, and neural network may be employed, among others.

An icon generating unit 24 generates an icon image as a special shape to be shown in the display, on the basis of the result of identifying the hand shape by the shape identifying unit 223. For example, when the result of identifying the shape of the hand was one finger, an icon of numeral "1" is generated as shown in Fig. 6 (A), or in the case of two fingers, an icon of numeral "2" is created as in Fig. 6 (B). As the shape of the icon, alternatively, when the result of identifying the shape of the hand was one finger, an icon of one finger may be shown as shown in Fig. 6 (C), or in the case of two fingers, an icon of two fingers may be created as shown in Fig. 6 (D). A display controller 23 controls the display on the basis of the result of identifying the shape of the hand by the shape identifying unit 223. For example, while displaying the icon according to the result of identifying, the menu previously corresponding to the result of identifying is displayed by emphasis on the basis of the hand shape identifying result.

In this embodiment of the invention, an example of operations described below. As shown in Fig. 7 (A), when the user confronts the appliance having the interface apparatus of the invention and points out one finger, an icon of numeral "1" is shown on the display, and the display of television on the first menu is shown by emphasis. At this time, by using sound or voice from the display device in tune with the emphasis display, the attention of the operator may be attracted. Herein, by pointing out two fingers as in Fig. 7 (B), an icon of numeral "2" is shown on the display and the display of network on the second menu is shown by emphasis. In this state, by maintaining the same hand shape for a specific time, the second menu is selected, and an instruction is given to the host computer so as to display the network terminal. For selection of menu, sound or the like may be used at the same time. In the case of hand shape different from those determined preliminarily as in Fig. 7 (C), icon and menu are not shown on the display, and no instruction is given to the host computer.

Thus, according to the invention, by identifying the shape of the hand in the taken image, it is possible to control the computer or appliance on the basis of the result of identifying, and it is possible to manipulate without making contact from a remote distance without using keyboard, mouse or other device. Besides, as the result of identifying the shape of the hand is reflected in the screen, the user can manipulate while confirming the result of identifying, and ease and secure manipulation is possible.

In this embodiment, this is an example of applying in selection of menu, but by pressing so that the icon display according to a specific shape of hand may be replaced by picture or message, it is also possible to control display and writing of picture or message.

(Second embodiment)

A second embodiment of the invention relates to an interface apparatus comprising a frame memory composed at least of an image pickup unit, a motion recognizing unit for recognizing the shape or move of an object in a taken picture, and a display unit for displaying the shape or move of the object recognized by the motion recognizing unit, for storing the image taken by the image pickup unit, and a reference image memory for storing the image taken before the image saved in the frame memory as reference image, wherein the motion recognizing unit comprises an image change depicting unit for depicting the difference between the image in the frame memory and the reference image stored in the reference image memory.

Fig. 8 shows the appearance of the second embodiment of the interface apparatus of the invention. In Fig. 8, same constituent elements as in the first embodiment are identified with same reference numerals. That is, reference 1 is a host computer, 2 is a display unit, and 3 is a CCD camera for picking up an image. The CCD camera 3 has an image pickup surface located in the same direction as the display direction, so that the hand gesture of the user can be picked up as the user confronts the display surface. On the display surface of the display unit 2, virtual switches 204, 205, 206, and an icon of an arrow cursor 203 for selecting the virtual switches are displayed.

Fig. 9 is a block diagram showing a specific constitution of the embodiment. The image fed through the CCD camera 3 is saved in a frame memory 21. A preliminarily taken image is stored in a reference image memory 25 as a reference image. A reference image updating unit 26 is composed of a timer 261 and an image updating unit 262, and is designed to update the reference image by transferring the latest image stored in the frame memory 21 to the reference image memory 25 at a specific time interval indicated by the timer 261.

An motion recognizing unit 22 depicts the difference between the image saved in the frame memory and the image stored in the reference image memory, and eliminates the background image from the image, and depicts the portion corresponding, for example, to the hand of the user, and also judges if the shape is one finger as shown in Fig. 10 (A) or a fist as shown in Fig. 10 (B).

Fig. 11 shows a detailed example of the motion recognizing unit 22, being composed of an image difference operation unit 221, a contour depicting unit 222, a shape change identifying unit 225, and a position detector 224.

The image difference operation unit 221 calculates the difference between the image saved in the frame memory 21 and the image stored in the reference image memory 25 as mentioned above. Consequently, the object desired to be depicted as motion, for example, the hand portion of the user, can be separated from the background portion, and only the moving object image can be depicted at the same time. For example, when the image difference operation unit 221 is composed of a mere subtraction circuit, as shown in Fig. 12, the hand portion in the reference image and only the hand portion of the latest image in the frame memory can be depicted, so that only the moving hand portion can be easily identified. The contour depicting portion 222 depicts the object existing in the image as the result of operation by the image difference operation unit 221, that is, the contour shape of the hand portion before moving and after moving. As an example of practical method, by depicting the edge of the image, the contour shape can be easily depicted.

The shape change identifying unit 225 identifies the detail of the contour shape of the hand portion after moving being depicted by the contour depicting unit 222, and judges if the shape is, for example, a finger as shown in Fig. 10 (A), or a fist as shown in Fig. 10 (B). At the same time, the position detector 224 calculates the coordinates of the center of gravity of the contour shape of the hand portion of the user after moving.

An icon generating unit 24 generates an icon image to be shown on the display on the basis of the result of identifying the hand shape by the shape change identifying unit 225. As examples of icon image, for example, when the result of identifying the hand shape is one finger, for example, the arrow marked icon as shown in Fig. 13 (A) may be generated, or in the case of a first shape, the x-marked icon as shown in Fig. 13 (B) may be generated. Or, if the identifying result of hand shape is two fingers, an icon mimicking two fingers as shown, for example, in Fig. 13 (C) may be generated, or in the case of a first, an icon mimicking the fist may be generated as shown in Fig. 13 (D).

A display controller 23 controls the display position of the icon generated by the icon generating unit 24 on the display 2, and is composed of a coordinate transforming unit 231 and a coordinate inverting unit 232. The coordinate transforming unit 231 transforms from the coordinates of the taken image into display coordinates on the display 2, and the coordinate inverting unit 232 inverts the lateral positions of the transformed display coordinates. That is, the coordinates of the center of gravity in the image of the portion corresponding to the user's hand detected by the position detector 224 are transformed into display coordinates on the display 2, and the lateral coordinates are inverted to display an icon in the display 2. By this manipulation, when the user moves the hand to the right, the icon moves to the right on the display screen, like a mirror action.

In thus constituted embodiment, an example of

operation is described below. As shown in Fig. 8, when the user confronts the appliance incorporating the interface apparatus of the embodiment and points out one finger of the hand, the arrow cursor appearing on the display moves to an arbitrary position corresponding to the move of the hand. Then, by moving the hand to an arbitrary one of the virtual switches 204, 205, 206 shown on the display 2, the arrow cursor is moved, and when the hand is gripped to form a fist, the one of the virtual switches 204, 205, 206 is selected, and an instruction is given to the host computer 1.

In this embodiment, it is designed to recognize the shape and move of the object in the taken image, but it may be also designed to recognize either the shape or the move of the object in the taken image.

Thus, according to the invention, which comprises the motion recognizing unit for recognizing the shape and/or move of the object in the taken image, display unit for displaying the shape and/or move of the object recognized by the motion recognizing unit, frame memory for saving the image taken by the image pickup means, and a reference image memory for storing the image taken before the image saved in the frame memory as reference image, by depicting the difference between the image in the frame memory and the reference image stored in the reference image memory in the motion recognizing unit, when the user confronts the image pickup unit and gives instruction by, for example, a hand gesture, the given hand gesture is shown on the display screen, and a virtual switch or the like shown on the display screen can be selected, for example, by the hand gesture, and a very simple manipulation of appliance without requiring input device such as mouse is realized.

(Third embodiment)

A third embodiment of the invention relates to an interface apparatus comprising contour depicting means composed of at least an image pickup unit, a motion recognizing unit for recognizing the shape and/or move of the hand of the user in the taken image, and a display unit for displaying the shape and/or move of the hand of the user recognized by the motion recognizing unit, thereby depicting the contour of the taken user image, a contour waveform operation unit for tracing the depicted contour, and calculating the relation between the angle of the contour line and the length of contour line, that is, the contour waveform, and a shape filter for filtering the contour waveform calculated by the contour waveform operation unit for generating a shape waveform expressing a specific shape, whereby composing the motion recognizing unit.

When the user confronts the image pickup unit of thus constituted interface apparatus and instructs by hand gesture, the image pickup unit picks up the user's image. The contour depicting means depicts the contour of the user's image, and the contour is transformed

into an angle of the contour line corresponding to the horizontal line, that is, as a contour waveform, by the contour waveform operation unit, on the horizontal axis in the length of the contour line starting from the reference point on the contour. This contour waveform is transformed into a shape waveform expressing the uneven shape of the finger by a shape filter composed of a band pass filter in specified band, for example, a band pass filter corresponding to uneven surface of finger, and the location of the finger is calculated, and only by counting the number of pulses existing in this shape waveform, the number of projected fingers, that is, the shape of the hand can be accurately judged. On the basis of the position or shape of the hand, the given hand gesture is shown on the display screen, and, for example, a virtual switch shown on the display screen can be selected by the hand gesture, and therefore very simple manipulation of appliance is realized without requiring mouse or other input device.

Moreover, plural shape filters may be composed of plural band pass filters differing in band, and the motion of the user may be judged on the basis of the shape waveforms generated by the plural shape filters. As a result, plural shapes can be recognized.

Alternatively, plural shape filters may be composed of at least of a band pass filter in the contour waveform shape corresponding to undulations of hand, and a band pass filter in the contour waveform shape corresponding to undulations of fingers. As a result, the image is transformed into a smooth shape waveform reflecting only undulations of the hand portion or into the shape waveform reflecting only undulations of the finger.

The motion recognizing unit may be constituted by comprising coordinate table for storing the contrast of coordinates of the contour shape of the taken image of the user and the contour shape calculated by the contour shape operation unit, and a coordinate operation unit for calculating the coordinates of the location of the specified shape in the taken image, by using the wave crest location position of the shape waveform and the coordinate table. Hence, the coordinates of the contour shape are calculated, and the coordinates are issued.

The motion recognizing unit may be also composed by comprising a shape judging unit for counting the number of pulses in the shape waveform generated by the shape filter, and the shape of the object may be judged by the output value of the shape judging unit. It is hence easy to judge whether hand is projecting two fingers or gripped to form a fist, by the number of pulses.

Also the motion recognizing unit may be composed by comprising a differentiating device for differentiating the shape waveform generated by the shape filter. By differentiation, the waveform is more pulse-like, and it is easier to count the number of pulses.

The appearance of the embodiment of the interface apparatus of the invention is similar to the one shown in

Fig. 8 relating to the second embodiment, and same parts as in the second embodiment are explained by referring to Fig. 8 and Fig. 10, and only other different parts are explained in Fig. 14 and after.

Fig. 14 is a detailed block diagram of the third embodiment of the invention. An image fed from the CCD camera 3 is stored in a frame memory 31. The motion recognizing unit 32 depicts the portion corresponding, for example, to the hand of the user from the image stored in the frame memory 31, and judges if the shape is, for example, one finger as shown in Fig. 10 (A), or a fist as shown in Fig. 10 (B).

Fig. 15 shows a detail of execution of the motion recognizing unit 32, and its detailed operation is described while referring also to Fig. 16 to Fig. 20.

A contour depicting unit 321 depicts the contour shape of the object existing in the image. As an example of specific method, the image is transformed into binary data, and by depicting the edge, the contour shape can be easily depicted. Fig. 17 (A1) is an example of depicted contour line, showing the hand projecting one finger.

A contour shape operation unit 322, starting from start point s in the diagram of the contour shape of the object depicted by the contour depicting unit 321 as shown in Fig. 17 (A1), traces the contour line in the direction of arrow in the drawing (counterclockwise), depicts the angle θ from the horizontal line of the contour line at each point x on the contour line as shown in Fig. 19 as the function in terms of the distance l from the start points, and transforms into the waveform shape regarding the distance l as the time axis as shown in Fig. 17 (B1), and simultaneously stores the coordinates of each point on the contour line corresponding to the distance l in a transformation table 324 in a table form. A shape filter 1 and a shape filter corresponding to reference numerals 3231 and 3232 respectively are filters for passing the band corresponding to the undulations of the hand portion and undulations of the finger portion, in the contour waveform shown in Fig. 17 (B1).

By the shape filter 1, Fig. 17 (B1) is transformed into a smooth shape waveform reflecting only the undulations of the hand portion as shown in Fig. 17 (B11), and by the shape filter 2, it is transformed into the shape waveform reflecting only the undulations of the finger as shown in Fig. 17 (B12), and both are differentiated by differentiating devices 3251 and 3252, and finally differential waveforms as shown in Fig. 17 (B112) and Fig. 17 (B122) are obtained. The shape judging unit 3262 judges whether the contour shape of the hand portion is two fingers as shown in Fig. 10 (A) or a fist as shown in Fig. 10 (B), and at the same time the coordinate operation unit 3261 calculates the coordinates of the center of gravity of the contour shape of the portion of the hand of the user. The coordinate operation unit 3261 determines positions $1c1$, $1c2$ of location of large pulse waveforms in the shape differential waveform shown in Fig. 17 (B112), and transforms into point $c1$ and point $c2$

shown in Fig. 20 by the coordinate transformation table 324, and the center of gravity of the hand portion is calculated from the contourline of the hand portion from point $c1$ to point $c2$, and issued as hand coordinates.

The shape judging unit 3262 counts and issues the number of pulse waveforms corresponding to the finger portion in the shape differential waveform in Fig. 17 (B122). That is, in the case of Fig. 17 (B122), since there are two large pulse waveforms corresponding to the portion of the finger, it is judged and issued as the shape of two fingers as shown in Fig. 10 (A). Or when the hand is gripped as shown in Fig. 18 (A2), there is almost no undulation of finger portion, and the output of the shape filter 2 is a shape waveform without undulations as shown in Fig. 18 (B22), and hence the output of the differentiating device 3262 is also a shape differential waveform without pulse waveform as shown in Fig. 18 (B222), and the count of pulse waveforms is 0, and it is hence judged and issued as the fist shape as shown in Fig. 10 (B).

As a practical example of composition of the shape judging unit 3262, simple threshold processing method or neural network maybe employed.

An icon generating unit 34 in Fig. 14 generates an icon image to be shown on the display on the basis of the result of identifying the shape of the hand by the shape judging means 3262 in Fig. 15. For example, if the result of identifying the hand shape is a shape of one finger, for example, an icon indicated by arrow shown in Fig. 16 (A) is created, or in the case of a fist shape, an icon indicated by x mark as shown in Fig. 16 (B) is created. A display controller 33 controls the display position of the icon generated by the icon generating unit 34 on the display, and is composed of coordinate transforming unit 331 and coordinate inverting unit 332. The coordinate transforming unit 331 transforms from the coordinates of the taken image into display coordinates on the display, and the coordinate inverting unit 332 inverts the lateral positions of the transformed display coordinates. That is, the coordinates of the center of gravity in the image of the portion corresponding to the user's hand issued by the coordinate operation unit 3261 in Fig. 15 are transformed into display coordinates on the display, and the lateral coordinates are inverted to display an icon in the display. By this manipulation, when the user moves the hand to the right, the icon moves to the right on the display screen, like a mirror action.

In thus constituted embodiment, an example of operation is described below. When the user confronts the appliance incorporating the interface apparatus of the embodiment and points out one finger of the hand, the arrow cursor of the icon 203 appearing on the display 2 moves to an arbitrary position corresponding to the move of the hand. Then, by moving the hand to an arbitrary one of the virtual switches 204, 205, 206 shown on the display 2, the arrow cursor is moved, and when the hand is gripped to form a fist, the one of the

virtual switches is selected, and an instruction is given to the host computer 1.

As an example of the icon to be displayed, as shown in Fig. 16 (C), (D), when the shape of the hand itself is formed into an icon, it corresponds to the move of the actual hand, and it is intuitive. More specifically, the images as shown in Fig. 16 (C) and (D) may be entered beforehand, or the contour shape data of the hand depicted by the contour depicting unit may be contracted or magnified to a desired size and used as an icon image.

Thus, in this embodiment, when the user confronts the image pickup unit of the interface apparatus and instructs, for example, by a hand gesture, the image pickup unit picks up the image of the user, and depicts the contour of the user's image, and transforms into an angle of contour line to the horizontal line, that is, into contour waveform, on the horizontal axis in the length of the contour line starting from the reference point on the contour line. This contour shape is transformed into a shape waveform expressing the uneven shape of the fingers by the shape filter composed of a band pass filter of specified band, for example, a band pass filter corresponding to undulations of fingers, and the position of the hand is calculated, and simultaneously the number of pulses existing in the shape waveform is counted, so that the number of projected fingers, that is, the shape of the shape can be accurately judged. On the basis of the position and shape of the hand, the given hand gesture is shown on the display screen, and, for example, a virtual switch shown on the display switch can be selected, and very easy manipulation of appliance is realized without requiring mouse or other input device.

(Fourth embodiment)

The foregoing embodiments relate to examples of manipulation on two-dimensional images shown on the display screen, whereas this embodiment relates to manipulation of a virtual three-dimensional image shown on a two-dimensional display screen.

Generally, assuming to grasp a virtual object in a virtual space by using a cursor, in a displayed virtual three-dimensional space, the following constitution is considered.

In Fig. 21, reference numeral A1 is an input device, A2 is a cursor coordinate memory unit, A3 is an object coordinate memory unit, A4 is a display device, and A5 is a contact judging unit. Fig. 22 (a) and Fig. 22 (b) show cursors in two-finger manipulator shape that can be expressed from the shape of the hand of the operator same as in the foregoing embodiments. Fig. 22 (a) shows an open finger state, and Fig. 22 (b) a closed finger state. Fig. 23 shows an example of a virtual object in a virtual space. Suppose the operator acts to grab the virtual object in a virtual three-dimensional space by using a two-finger cursor. Fig. 24 (a) and Fig. 24 (b) show configuration of cursor and virtual object in the vir-

tual space when gripping the virtual object by using the cursor. Fig. 25 shows the display of the display device A4.

When manipulation of the operator is given to the input unit A1, the cursor coordinates and the two-finger interval of the cursor stored in the cursor coordinate memory device A2 are updated according to the manipulation. The display device A4 depicts the virtual space including the cursor and virtual object by using the information stored in the cursor coordinate memory unit A2 and the position information of the virtual object stored in the object coordinate memory unit A3. Herein, the contact judging unit A5 calculates whether the cursor and virtual object contact with each other in the virtual space or not, by using the information stored in the cursor coordinate memory unit A2 and the position information of the virtual object stored in the object coordinate memory unit A3. More specifically, the distance between plural surfaces composing the cursor and virtual object in the virtual space is calculated on each surface, and when the virtual object contacts between two fingers of cursor, it is judged that the cursor has grabbed the virtual object, and thereafter the coordinates of the object are changed according to the move of the cursor.

In such technique, however, the display by the display device is as shown in Fig. 25 in the case of configuration as shown in Fig. 24 (a) or (b), and the operator may misjudge that the coordinates are matched although the cursor and virtual object position are not matched exactly in the virtual space. Or, in the case of using the three-dimensional display device or in the case of combined display of Fig. 24 (a) and (b), smooth manipulation is difficult due to difference in the sense of distance in the actual space and in the virtual space.

Thus, due to difference between the sense of distance in a virtual space which is a display space and the sense of distance in an actual space, or due to difference between the motion of cursor intended by the operator and the actual motion of cursor, interaction according to the intent of the operator (in this case, grabbing of the virtual object) cannot be realized smoothly in the interaction of the cursor and virtual object in virtual space (for example, when grabbing the virtual object by a virtual manipulator).

In this embodiment, the cursor can be controlled with ease by hand gesture or the like by the operator in the virtual space without making contact, and presence or absence of occurrence of interaction with the virtual object in the virtual space is determined not only by the distance between the cursor in the virtual space and the constituent element of the virtual space (the surface in the case of three-dimensional virtual space), but by inducing interaction also on the object of which distance in the virtual space is not necessarily close by the interaction judging means, the judgment whether the cursor induces interaction with the virtual object is made closer to the intent of the operator, so that the controllability of

the interface may be further enhanced. It is further possible to induce interaction also on the object of which distance in the virtual space is not necessarily close.

A first constitution of the embodiment is an interface apparatus comprising display means, input means for changing the position and shape of the cursor displayed in the display means, cursor memory means for storing coordinates of a representative point representing the position of the cursor and the shape of the cursor, object memory means for storing coordinates of a representative point representing the position of display object other than the cursor and shape of the display object, and interaction judging means for judging interaction between the cursor and the display object, by using the position and shape of the cursor stored in the cursor memory means and the position and shape of the display object stored in the object memory means, wherein the interaction judging means is composed of distance calculating means for calculating the distance between at least one representative point of the cursor and at least one representative point of the display object, motion recognizing means for recognizing the move of the cursor or change of the shape, and overall judging means for determining the interaction of the cursor and display object by using the distance calculated by the distance calculating means and the result of recognition by the motion recognizing means.

According to this constitution, presence or absence of occurrence of interaction between the cursor manipulated by the operator in the virtual space and the virtual object in the virtual space is determined not only by the distance between the cursor in the virtual space and the constituent element of the virtual object (the surface in the case of a three-dimensional virtual space), but the overall judging means judges the presence or absence of occurrence of interaction by the distance between representative points calculated by the distance calculating means and the motion of the cursor recognized by the motion recognizing means, so that interaction may be induced also on the object of which distance is not necessarily close in the virtual space.

When the motion recognizing means recognizes a preliminarily registered motion in the first constitution, a second constitution may be designed so that the interaction judging means may induce an interaction on the display object of which distance calculated by the distance calculating means is below a predetermined reference.

In the first and second constitutions, by installing move vector calculating means for calculating the moving direction and moving distance of the cursor in the display space to compose the interaction judging means, a third constitution may be composed so as to determine the interaction of the cursor and display object by the interaction judging means on the basis of the moving direction of the cursor and moving distance of the cursor calculated by the move direction calculating means.

The third constitution may be modified into a fourth constitution in which the interaction judging means generates an interaction when the cursor moving distance calculated by the move vector calculating means is less than the predetermined reference value.

In third and fourth constitutions, the interaction judging means may generate an interaction on the display object existing near the extension line in the moving direction of the cursor calculated by the move vector calculating means, so that a fifth constitution may be composed.

In the first to fifth constitutions, the interaction judging means may generate an interaction when the shape of the cursor and shape of the display object become a preliminarily registered combination, which may be composed as a sixth constitution.

In the first to sixth constitutions, by composing the interaction judging means by incorporating shape judging means for recognizing the shape of the cursor and shape of the display object, a seventh constitution may be constructed so that the interaction judging means may generate an interaction when the shape of the cursor and shape of the display object recognized by the shape recognizing means coincide with each other.

In the first to seventh constitutions, by comprising sight line input means for detecting sight light direction, an eighth constitution may be composed in which the interaction judging means generates an interaction when the motion recognizing means recognizes a preliminarily registered motion, on the display object near the extension line of the sight light detected by the sight line input means.

The eighth constitution may be modified into a ninth constitution in which the interaction judging means generates an interaction when the cursor is present near the extension line of the sight line on the display object near the extension line of the sight line detected by the sight line input means and the motion recognizing means recognizes a preliminarily registered motion.

In the first to ninth constitutions, when an interaction is generated, learning means may be provided for learning the configuration of the cursor and the objective display object, and the shape of the cursor and shape of the display object, so that a tenth constitution may be composed for determining the interaction on the basis of the learning result of the learning means by the interaction judging means.

The tenth constitution may be modified into an eleventh constitution in which the interaction judging means generates an interaction when the configuration of the cursor and objective display object, or the shape of the cursor and shape of the display object may be similar to the configuration or shapes learned in the past by the learning means.

Instead of the first to eleventh constitutions, a twelfth embodiment may be composed in which the interaction judging means may be composed by incorporating coordinate transforming means for transform-

ing the coordinates from the cursor memory unit and object memory unit to the input to the distance calculating means.

The twelfth constitution may be modified into a thirteenth constitution in which the cursor and objective display object may be brought closer to each other when an interaction is generated.

The fourth embodiment is described in detail by referring to the drawing. Fig. 26 is a block diagram of the interface apparatus of the embodiment.

In Fig. 26, reference numeral 41 is input means, 42 is cursor memory means, 43 is object memory means, 44 is display means, 45 is interaction judging means, 45a is distance calculating means, 45b is motion recognizing means, 45c is overall judging means, 45d is move vector calculating means, 45e is shape judging means, 45f is learning means, 45g is coordinate transforming means, and 46 is sight line input means.

In Fig. 26, the operator manipulates input means 41, the cursor memory means 42 changes and stores the coordinates and shape of representative point representing the position in the virtual space of the cursor, and the display means 44 shows the cursor and virtual object in two-dimensional display or three-dimensional display on the basis of the coordinates and shape of the representative point representing the position in the virtual space of the cursor stored in the cursor memory means 42 and the coordinates and shape of representative point representing the position in the virtual space of the virtual object stored in the object memory means 43.

The sight line input means 46 detects the position of the sight line of the operator on the display. The distance calculating means 45a calculates the distance between the cursor and virtual object in the virtual space on the basis of the coordinates of the representative points stored in the cursor memory means 42 and object memory means 43. The motion recognizing means 45b recognizes the motion of manipulation on the basis of the data stored in the cursor memory means 42 and object memory means 43. The move vector calculating means 45d calculates the moving direction and moving distance of the cursor in the virtual space. The shape judging means 45e judges whether the shape of the cursor and shape of the virtual object are appropriate for inducing an interaction or not. The learning means 45f stores the relation of position and shape of the cursor and virtual object when the overall judging means 45c has induced an interaction between the cursor and virtual object, and tells whether the present state is similar to the past state of inducing interaction or not.

The overall judging means 45c judges whether the cursor and virtual object interact with each other or not, on the basis of the distance between the cursor and virtual object issued by the distance calculating means 45a, the result recognized by the motion recognizing means 45b, moving direction and moving distance of

cursor calculated by the move vector calculating means 45d, the position of sight line detected by the sight light input means 46, judging result of the shape judging means 45e, and degree of similarity to the past interaction issued by the learning means 45f, and changes the coordinates and shape of the representative points of the cursor and virtual object depending on the result of interaction. The coordinate transforming means 45g transforms the coordinates of the cursor and objective virtual object in the virtual space used in the distance calculation by the distance calculating means 45a so that the positions of the two may be closer to each other when the interaction judging means 45 induces an interaction.

Fig. 22 (a) and (b) show a two-finger manipulator shape in a first example of the cursor used in the interface apparatus of the invention. In the diagram, the fingers are opened in (a) and the fingers are closed in (b). Fig. 22 (c) and (d) show a two-finger two-joint manipulator shape in a second example of the cursor used in the interface apparatus of the invention. In the diagram, the fingers are opened in (c) and the fingers are closed in (d). Fig. 22 (e) and (f) show a five-finger hand shape in a third example of the cursor used in the interface apparatus of the invention. In Fig. 22, the hand is opened in (e) and the hand is closed in (f).

Fig. 23 (a) and (b) show examples of the object in the virtual space used in the interface apparatus of the invention, showing a cube in (a) and a plane in (b).

In the interface apparatus thus constituted, the operation is described below. In this embodiment, it is supposed that the operator moves the cursor as shown in Fig. 22 in a three-dimensional virtual space, and moves by grabbing the virtual object as shown in Fig. 23 existing in the virtual space.

Manipulation of the operator is effected on the input means 41. In the input herein, as the input device for feeding information for varying the position or shape of the cursor, the means as shown in Fig. 27 (a) to (c), or the camera, keyboard, or command input by voice recognition can be used. Fig. 27 (a) relates to a mouse, and the cursor is manipulated by moving the mouse main body or clicking its button. Fig. 27 (b) relates to a data glove, which is worn on the hand of the operator, and the cursor is manipulated by reflecting the finger joint angle or position of the data glove in the actual space in the position and shape of the cursor. Fig. 27 (c) relates to a joy stick, and the cursor is manipulated by combination of lever handling and operation button. When using a camera, the body or part of the body (for example, the hand) is taken by the camera, and the shape and position of the hand are read.

Fig. 28 shows an example of shape depiction when only the hand is taken by the camera. In Fig. 28 (a), the hand is taken by the camera. The luminance of pixels of the image in Fig. 28 (a) is converted into binary data in Fig. 28 (b). In Fig. 28 (b), it is possible to judge the degree of opening or closing of the hand by the ratio of

the longer side and shorter side of a rectangle circumscribing a black region, and input of position and distance is enabled from the coordinates of center of gravity and area of the entire black pixels. The input means 41 sends the manipulation data (cursor moving distance, cursor shape changing amount, etc.) to the cursor memory means 42.

The cursor memory means 42 stores the coordinates and shape of the representative point of the cursor in the virtual space stored in the cursor memory means on the basis of the manipulation data sent out by the input means 41. As the representative point, the coordinates of center of gravity of cursor (X_0, Y_0, Z_0) may be used. Moreover, as the representative point, the coordinates of the center of each surface composing the cursor or the coordinates of apex may be also used. As the shape, the two-finger interval d in the case of Fig. 22 (a), or the internal angle θ_n of the joint of each finger in the case of Fig. 22 (b), (c) (n is a joint number: as θ_n becomes smaller, the joint is bent) may be used as the storage information. Moreover, as the shape, the finger tip of each finger or the coordinates of each joint in the virtual space may be also used.

The object memory means 43 stores the coordinates and shape of the representative point of the virtual object in the virtual space shown in Fig. 23 as the object of manipulation. As the representative point, the coordinates of the center of gravity of virtual object (cube: (X_1, Y_1, Z_1), plane: (X_2, Y_2, Z_2)) are used. Also as the representative point, the coordinates of the center of each surface composing the virtual object or the coordinates of the apex may be used. As the shape, parameter α showing a predetermined shape is stored (herein cube is defined as $\alpha=1$, and plane as $\alpha=2$). Also as the shape, the coordinates of apex may be used.

The display means 44 shows the image in two-dimensional display as seen from the viewpoint preliminarily assuming the virtual space on the basis of the information of the position and shape of the cursor and virtual object stored in the cursor memory means 42 and object memory means 43. Fig. 29 (a) shows a display example of display means. When the operator manipulates, the display position or shape of the cursor is changed, and the operator continues to manipulate according to the display.

The interaction judging means 45 judges if the cursor has grabbed the object or not (presence or absence of interaction) every time the cursor position changes, and when it is judged that the cursor has grabbed the object, the coordinates of the virtual object are moved according to the move of the cursor.

The distance calculating means 45a calculates the distance between the coordinates of the center of gravity of the cursor (X_0, Y_0, Z_0) stored in the cursor memory means 42 and the coordinates of the center of gravity of the virtual object (X_1, Y_1, Z_1), (X_2, Y_2, Z_2) stored in the object memory means 43.

The motion recognizing means 45b recognizes the

motion of "grab" as the preliminarily registered motion by using the change of shape of the cursor. In the case of the cursor in Fig. 22 (a), the decreasing state of the interval d of two fingers is recognized as the "grab" action, and in the case of the cursor in Fig. 22 (b), (c), the decreasing state of the angle θ_n of all fingers is understood as the "grab" action. As the technique of recognizing the motion, meanwhile, the time series changes of the parameters representing the shape (such as d and θ_n mentioned above) may be used as the recognizing technique after learning specific motions preliminarily by using the time series row pattern recognition techniques (table matching, DP matching, hidden Markoff model (HMM), recurrent neural network, etc.).

The move vector calculating means 45d calculates the moving direction and moving distance of the cursor in the virtual space by using the changes of the coordinates of the center of the cursor (X_0, Y_0, Z_0). For example, the direction and magnitude of the differential vector of the coordinates of the center of gravity of the present time t (X_0, Y_0, Z_0) $_t$ and the coordinates of the center of gravity of a certain previous time (X_0, Y_0, Z_0) $_{t-1}$ are used as moving distance of the cursor

The shape judging means 45e judges if the shape of the cursor for storing in the cursor memory means is proper or not for grabbing the virtual object in the shape stored in the object memory means (whether the cursor shape is appropriate or not for inducing interaction with the virtual object). Herein, when the value of the parameter α representing the shape of the object is 1, the cursor finger open state is regarded as an appropriate state. The cursor finger open state is judged, for example, when the value of d is larger than the intermediate value of maximum value of d , that is, d_{max} , and 0 in the case of the cursor shown in Fig. 22 (a), and when all joint angles θ_n are greater than the intermediate value of maximum value θ_n max and 0 in the case of Fig. 22(b), (c).

When the value of the parameter α expressing the object shape is 0, it is an appropriate state when the interval of finger tips of the cursor is narrow. The finger tips of the cursor are judged to be narrow when, for example, when the value of d is smaller than the intermediate value of the maximum value of d , d_{max} and 0 in the case of the cursor in Fig. 22 (a), or when all joint angles θ_n are smaller than the intermediate value of the maximum value θ_n max and 0 in the case of Fig. 22 (b), (c). As the judging method of the shape, incidentally, the parameter expressing the cursor shape (d or θ_n) is stored preliminarily in the state of the cursor grabbing the virtual object in contact state in the virtual space, and when the values of the parameters coincide in a range of $\pm 30\%$, it may be judged to be appropriate for grabbing action.

The sight line input means 46 detects the sight line of the operator, and calculates the coordinates noticed by the operator on the display screen of the display

means 44 (coordinates of notice point). As the sight light detecting means, by detecting the direction of the pupil of the operator by using a photo sensor such as CCD camera, the notice point on the screen is calculated by measuring the position of the head of the operator by using a camera or the like.

The learning means 45f stores the parameter (d or θ n) showing the shape of the cursor when the overall judging means 45c judges that the cursor has grabbed the virtual object, the parameter α showing the shape of the grabbed object, and the relative configuration of the position of the cursor and the position of the virtual object (the vector linking the center of gravity of the cursor and the center of gravity of the virtual object), and when the parameter expressing the present shape of the virtual object, the parameter expressing the shape of the surrounding virtual object, and the configuration of the present center of gravity of the cursor and the center of gravity of the surrounding virtual object are close to the past state of grabbing the object (for example, when the parameters and element values of each dimension of the vector expressing the configuration coincide with the past values within a range of $\pm 30\%$), it is judged to be close to the past situation and 1 is issued, and otherwise 0 is issued. As other means of learning, meanwhile, the parameter expressing the shape of the cursor when grabbing the object in the past, the parameter α expressing the shape of the grabbed virtual object, and the relative configuration of the position of the cursor and position of the virtual object maybe learned by using neural networks or the like. As the learning items, it may be also possible to learn together with the configuration of notice point coordinates on the screen detected by the sight line detecting means 46 and coordinates of the cursor on the display screen.

The coordinates transforming means 45g transforms the coordinates used in distance calculation by the distance calculating means when grabbing the object (when an interaction is caused) so that the distance between the cursor and the objective virtual object in the virtual space may be shorter. For example, supposing the coordinates to be (100, 150, 20) and (105, 145, 50) when the cursor grabs the virtual object, the coordinates transforming means transforms the Z-coordinate having the largest difference among the coordinates as shown in formula (1).

$$Z' = 0.8 \times Z \quad (1)$$

where z is the Z-coordinate of the center of gravity of the cursor and virtual object received by the coordinates transforming means, and Z' denotes the Z-coordinate as the output of the coordinates transforming means.

In this case, the value of X-coordinate and value of Y-coordinate are not changed. Also the values stored in the cursor memory means and object memory means are not changed, and hence the screen shown by the

display means is not changed. By such transformation, if the distance in the virtual space is remote, when the operator attempts to grab, thereafter the distance between the cursor and virtual object becomes shorter when calculating the distance, so that the distance calculating means can calculate the distance closer to the sense of the distance felt by the operator.

The overall judging means 45c judges the occurrence of interaction of "grab" when recognizing the preliminarily registered action of "grab" by the motion recognizing means 45b when the distance between the cursor and the virtual object issued by the distance calculating means 45a is less than a predetermined reference, and thereafter until the interaction of "grab" is terminated, the values of the coordinates of the center of gravity of the grabbed virtual object stored in the object memory means 43 are matched with the coordinates of the center of gravity of the cursor. Herein, the predetermined reference value may be a larger than the actual distance of contact of the cursor and object in the virtual space. For example, in the case of Fig. 25 (configuration of Fig. 24), if the distance between the virtual object and cursor is less than the reference value of the distance, the action of grabbing by the operator may be instructed to the input means 1, and when the motion recognizing means 45b recognizes the grabbing action, the virtual object may be grabbed and moved.

Meanwhile, the overall judging means 45c, if there are plural virtual objects below the reference of the distance, judges only the objects below the reference (for example, 90 degrees) of the angle formed by the line segment (waved line) linking the cursor and virtual object and the moving direction (arrow) of the cursor calculated by the move vector calculating means 45d as shown in Fig. 29 (b), so that the operator can judge the interaction in consideration of the moving direction of the cursor in the process of manipulation (selecting the highest position of the three objects in the diagram). As for the cursor moving distance, if the moving distance is longer than the predetermined moving distance reference, interaction does not occur. Thus, when merely moving the cursor, interaction not intended by the operator is not caused.

Moreover, as shown in Fig. 29 (c), when plural virtual objects are satisfying the reference, the virtual object close to the position of the notice point detected by the sight line input means 46 is the object of grabbing by the overall judging means 45c (in the diagram, the object at the left side close to the "+" mark indicating the notice point is to be selected). As a result, the object can be easily selected by using the sight line of the operator.

Incidentally, as shown in Fig. 29 (d), if there is a close object on the screen, the virtual object coinciding with the shape of the cursor judged by the shape judging means 45e is the object of grabbing by the overall judging means 45c (in the diagram, since the finger interval of the cursor is narrow, a plane is judged to be appropriate as the object of grabbing, and hence the

plane is selected). As a result, the virtual object intended by the operator can be selected by the shape of the cursor, and the operator can manipulate easily by corresponding to the cursor shape that can be easily associated when grabbing the virtual object.

The overall judging means 45c selects by priority the virtual object judged to be similar when the object was grabbed in the past by the learning means 45f. As a result, the judgment closer to the past manipulation by the operator can be reproduced, and the controllability may be enhanced.

Thus, according to the invention, presence or absence of interaction between the cursor manipulated by the operator in the virtual space and the virtual object in the virtual space is determined not only by the distance between the cursor and virtual object, but it is determined on the basis of the action, sight line or past case in the manipulation of the operator, so that the controllability may be enhanced in the interface for interaction with the virtual object by using the cursor in the virtual space.

The embodiments have been explained by referring to the action of grabbing the virtual object by using the cursor as the interaction, but similar handling is also possible in other motions, such as indicating (pointing) to the virtual object, collision, friction, impact, and remote control. Similar effects are obtained if the virtual space is a two-dimensional space or if the display means is a three-dimensional display. It may be realized by using hardware, or by using the software on the computer.

In this way, according to the embodiments, presence or absence of occurrence of interaction between the cursor manipulated by the operator in the virtual space and the virtual object in the virtual space is determined not only by the distance between the constituent elements of the cursor and virtual object in the virtual space, but the overall judging means judges presence or absence of occurrence of interaction on the basis of the distance between representative points calculated by the distance calculating means and the motion of the cursor recognized by the motion recognizing means, and therefore an interaction may be induced also on the object of which distance is not necessarily close in the virtual space, so that the input and output interface excellent in controllability may be presented. Moreover, it is not necessary to calculate the distance of all the constituent elements between the cursor and virtual object in the virtual space as required in the conventional contact judging method, and therefore the quantity of calculation is lessened, and the processing speed is enhanced.

Accordingly, the invention can recognize the shape or motion of the hand of the operator and display the feature of the shape of the recognized hand as a cursor on the screen as the special shape, so that the information displayed on the screen and the information displayed on the screen by the shape or motion of the hand

can be controlled easily at superior controllability.

Moreover, the feature of the shape of the hand is displayed as cursor on the screen as the special shape, and the relation with other display objects than the cursor is judged sequentially and automatically by the interaction along the intent of the operator, so that the interface further enhanced in the controllability of manipulation of indicating or gripping the display object can be realized.

EXPLANATION OF NOTATIONS OF THE DRAWINGS

1	Host computer
2	Display
3	CCD camera
21	Frame memory
22	Shape identifying means
23	Display control unit
24	Icon generating unit
25	Reference image memory
26	Reference image updating unit
31	Frame memory
32	Motion recognizing means
33	Display control unit
34	Icon generating unit
41	Input means
42	Cursor memory means
43	Object memory means
44	Display means
45	Interaction judging means
46	Sight line input means
200	Icon
201	Menu
202	Menu
203	Icon
204	Virtual switch
205	Virtual switch
206	Virtual switch
221	Image difference calculation unit
222	Contour depicting unit
223	Shape identifying unit
224	Position detector
225	Shape change identifying unit
231	Coordinate transforming unit
232	Coordinate inverting unit
261	Timer
262	Image updating unit
321	Contour depicting unit
322	Contour waveform operation unit
324	Coordinate table
331	Coordinate transforming unit
332	Coordinate inverting unit
501	Host computer
502	Display
503	Operation button
504	Operation button
505	Operation button
506	Mouse cursor

507 Mouse
 508 Switch
 3231 Shape filter 1
 3232 Shape filter 2
 3251 Differentiating device
 3252 Differentiating device
 3261 Coordinate operation unit
 3262 Shape judging unit(differentiating device)

Claims

1. An interface apparatus comprising:

recognizing means for recognizing the shape
 of a hand of an operator,
 display means for displaying the features of the
 shape of the hand recognized by the recognizing
 means on the screen as a special shape,
 and
 control means for controlling the information
 displayed in the screen by the special shape
 displayed in the screen by the display means.

2. An interface apparatus of claim 1, wherein the recognizing means recognizes the motion of the hand together with the shape of the hand, and the display means displays the features of the shape and motion of the hand recognized by the recognizing means on the screen as the special shape.

3. An interface apparatus of claim 1, wherein the control means controls so as to select the information displayed in the screen.

4. An interface apparatus comprising:

a frame memory composed of

at least an image pickup unit,
 a motion recognizing unit for recognizing
 the shape or move of an object in a taken
 picture, and
 a display unit for displaying the shape or
 move of the object recognized by the
 motion recognizing unit, for
 storing the image taken by the image
 pickup unit, and
 a reference image memory for storing the
 image taken before the image saved in the
 frame memory as reference image,
 wherein the motion recognizing unit
 comprises an image change depicting unit
 for depicting the difference between the
 image in the frame memory and the refer-
 ence image stored in the reference image
 memory.

5. An interface apparatus of claim 4, further compris-

ing a reference image updating unit for updating the reference image accumulated in the reference image memory into a novel image.

6. An interface apparatus of claim 4, further comprising a timer for calculating the interval of updating of reference image provided in the reference image updating unit.

7. An interface apparatus comprising contour depicting means composed of:

at least an image pickup unit,
 a motion recognizing unit for recognizing the
 shape and/or move of the hand of the user in
 the taken image, and
 a display unit for displaying the shape and/or
 move of the hand of the user recognized by the
 motion recognizing unit, thereby
 depicting the contour of the taken user image,
 a contour waveform operation unit for tracing
 the depicted contour, and calculating the rela-
 tion between the angle of the contour line and
 the length of contour line, that is, the contour
 waveform, and a
 shape filter for filtering the contour waveform
 calculated by the contour waveform operation
 unit for generating a shape waveform express-
 ing a specific shape, whereby composing the
 motion recognizing unit.

8. An interface apparatus of claim 7, wherein plural shape filters are composed by plural band pass filters differing in band, and the motion of the user is judged on the basis of the shape waveform generated by the plural shape filters.

9. An interface apparatus of claim 7, wherein plural shape filters are composed of at least a band pass filter of contour waveform shape corresponding to undulations of hand, and a band pass filter of contour waveform shape corresponding to undulations of fingers.

10. An interface apparatus of claim 7, wherein the motion recognizing unit is composed by comprising a coordinate table for storing the contrast of the coordinates of the contour shape of the taken user image and the contour waveform calculated by the contour shape operation unit, and a coordinate operation unit for calculating the coordinates of the location of the specified shape in the taken image by using the wave crest location of the shape waveform and the coordinate table.

11. An interface apparatus of claim 7, wherein the motion recognizing unit is composed by containing a shape judging unit for counting the number of

pulses in the shape waveform generated by the shape filter, and the shape of the object is judged by the output value of the shape judging unit.

12. An interface apparatus of claim 7, wherein the motion recognizing unit is composed by containing differentiating devices for differentiating the shape waveforms generated by the shape filter. 5
13. An interface apparatus comprising: 10
display means,
input means for changing the position and shape of the cursor displayed in the display means, 15
cursor memory means for storing coordinates of a representative point representing the position of the cursor and the shape of the cursor,
object memory means for storing coordinates of a representative point representing the position of display object other than the cursor and shape of the display object, and 20
interaction judging means for judging interaction between the cursor and the display object, by using the position and shape of the cursor stored in the cursor memory means and the position and shape of the display object stored in the object memory means, 25
wherein the interaction judging means is composed of 30
distance calculating means for calculating the distance between at least one representative point of the cursor and at least one representative point of the display object, 35
motion recognizing means for recognizing the move of the cursor or change of the shape, and
overall judging means for determining the interaction of the cursor and display object by using the distance calculated by the distance calculating means and the result of recognition by the motion recognizing means. 40
45
14. An interface apparatus of claim 13, wherein when the motion recognizing means recognizes a preliminarily registered motion, the interaction judging means induces an interaction on the display object of which distance calculated by the distance calculating means is below a predetermined reference. 50
15. An interface apparatus of claim 13, wherein by installing move vector calculating means for calculating the moving direction and moving distance of the cursor in the display space, the interaction judging means is composed, and the interaction judging 55

means determines the interaction of the cursor and display object by the interaction judging means on the basis of the moving direction of the cursor and moving distance of the cursor calculated by the move direction calculating means.

16. An interface apparatus of claim 15, wherein the interaction judging means generates an interaction when the cursor moving distance calculated by the move vector calculating means is less than the predetermined reference value.
17. An interface apparatus of claim 15, wherein the interaction judging means generates an interaction on the display object existing near the extension line in the moving direction of the cursor calculated by the move vector calculating means.
18. An interface apparatus of claim 13, wherein the interaction judging means generates an interaction when the shape of the cursor and shape of the display object become a preliminarily registered combination.
19. An interface apparatus of claim 13, wherein by composing the interaction judging means by incorporating shape judging means for recognizing the shape of the cursor and shape of the display object, the interaction judging means generates an interaction when the shape of the cursor and shape of the display object recognized by the shape recognizing means coincide with each other.
20. An interface apparatus of claim 13, wherein by comprising sight line input means for detecting sight light direction, the interaction judging means generates an interaction when the motion recognizing means recognizes a preliminarily registered motion, on the display object near the extension line of the sight light detected by the sight line input means.
21. An interface apparatus of claim 20, wherein the interaction judging means generates an interaction when the cursor is present near the extension line of the sight line on the display object near the extension line of the sight line detected by the sight line input means and the motion recognizing means recognizes a preliminarily registered motion.
22. An interface apparatus of claim 13, wherein when an interaction is generated, learning means may be provided for learning the configuration of the cursor and the objective display object, and the shape of the cursor and shape of the display object, so that the interaction judging means determines the interaction on the basis of the learning result of the learning means.

23. An interface apparatus of claim 22, wherein the interaction judging means generates an interaction when the configuration of the cursor and objective display object, or the shape of the cursor and shape of the display object may be similar to the configuration or shapes learned in the past by the learning means. 5
24. An interface apparatus of claim 13, wherein the interaction judging means may be composed by incorporating coordinate transforming means for transforming the coordinates from the cursor memory unit and object memory unit to the input to the distance calculating means. 10
25. An interface apparatus of claim 24, wherein the cursor and objective display object may be brought closer to each other when an interaction is generated. 15
- 20
- 25
- 30
- 35
- 40
- 45
- 50
- 55

Fig. 1

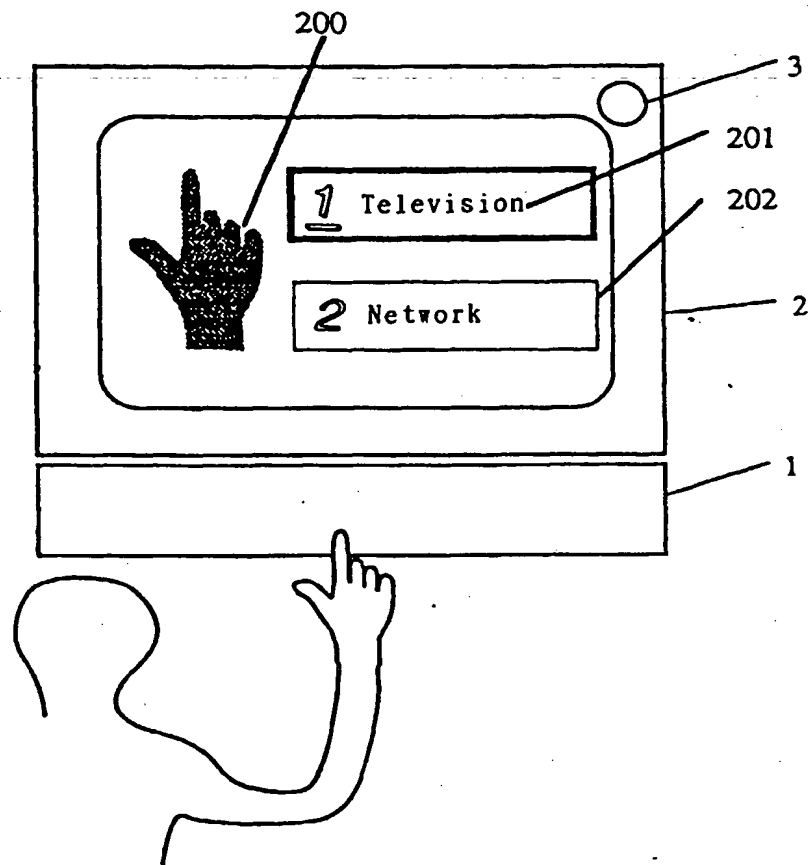


Fig. 2

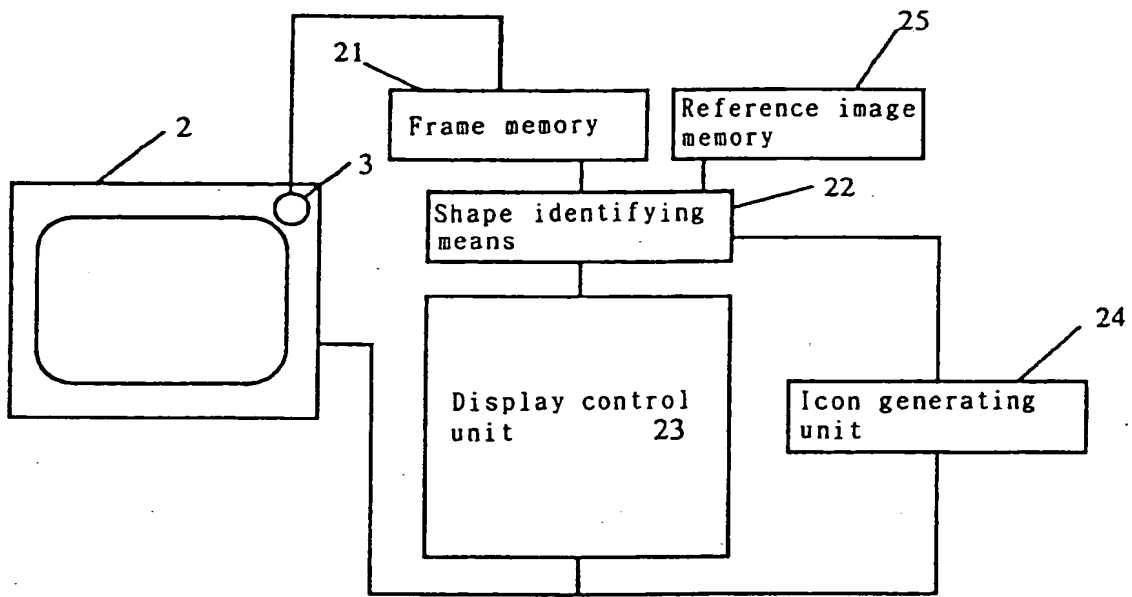


Fig. 3

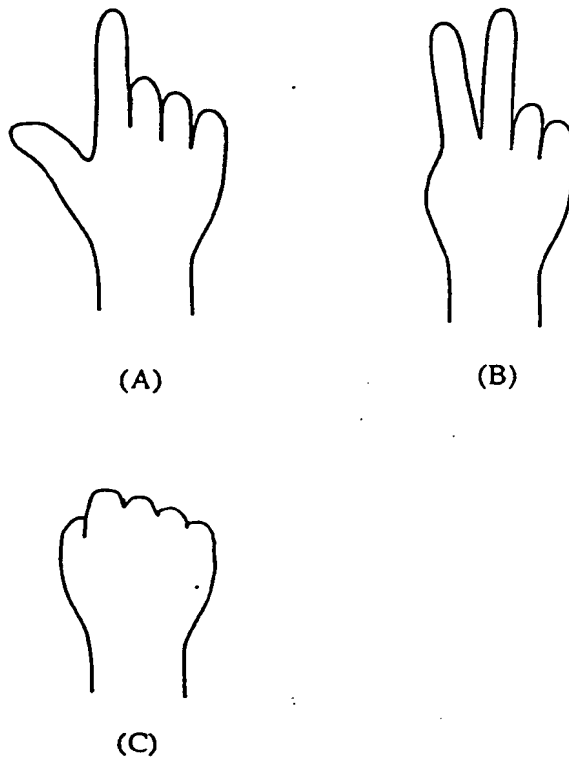


Fig. 4

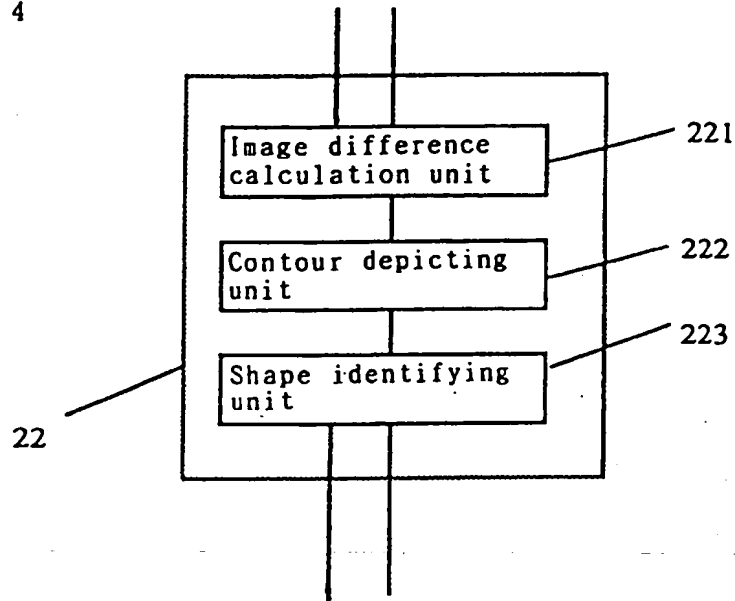


Fig. 5

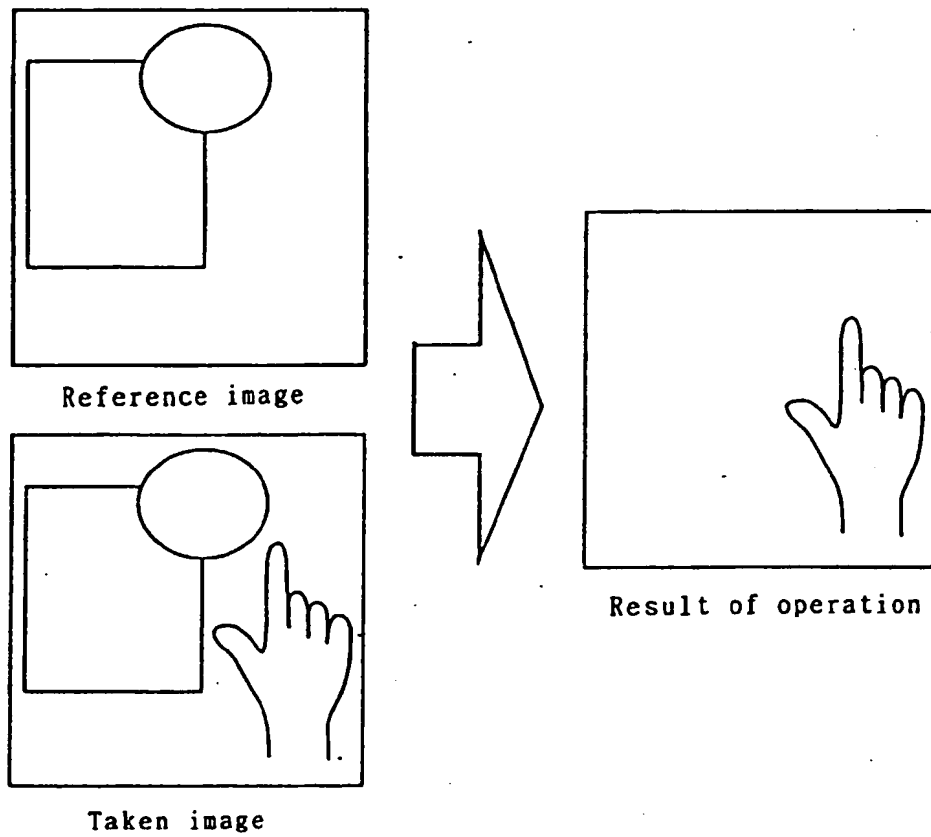


Fig. 6

1

(A)

2

(B)



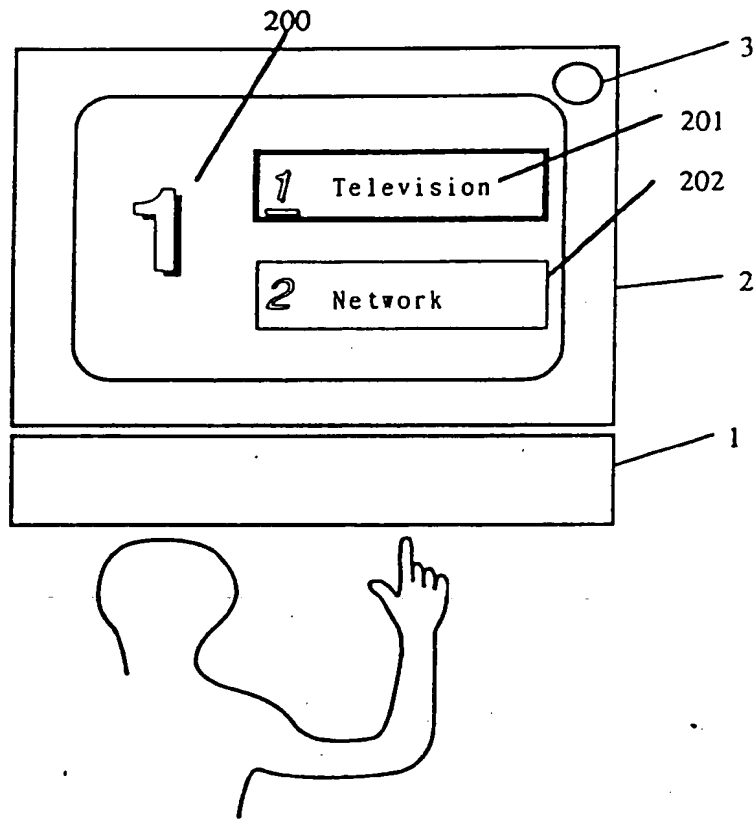
(C)



(D)

Fig. 7

(A)



(B)

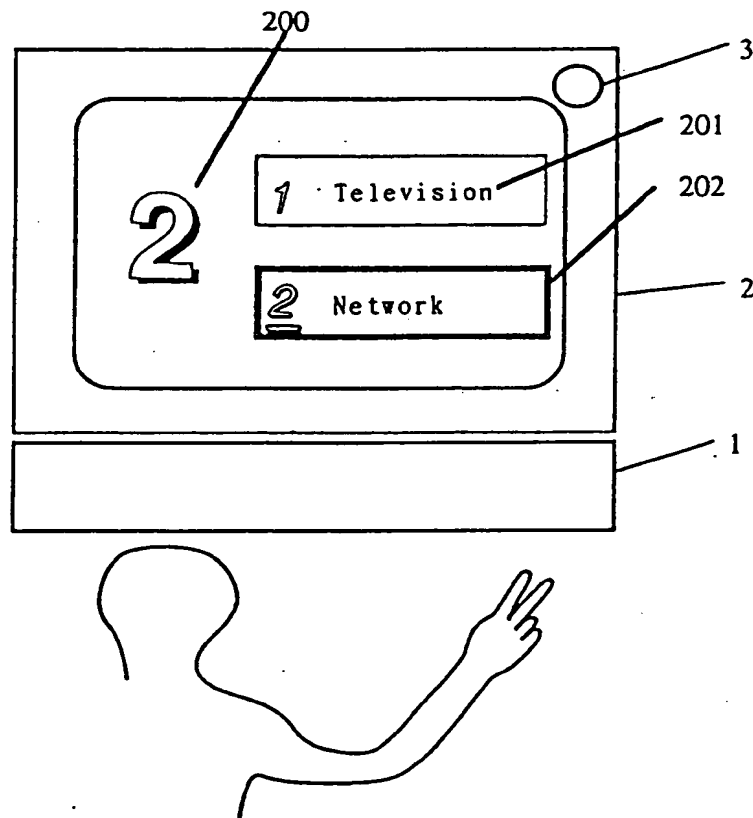


Fig. 7

(C)

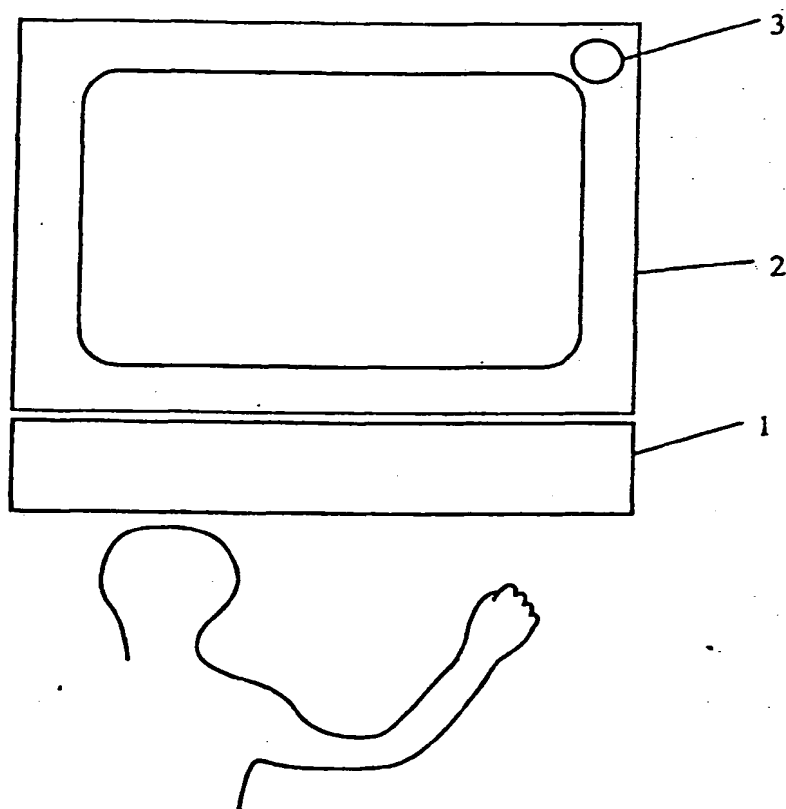
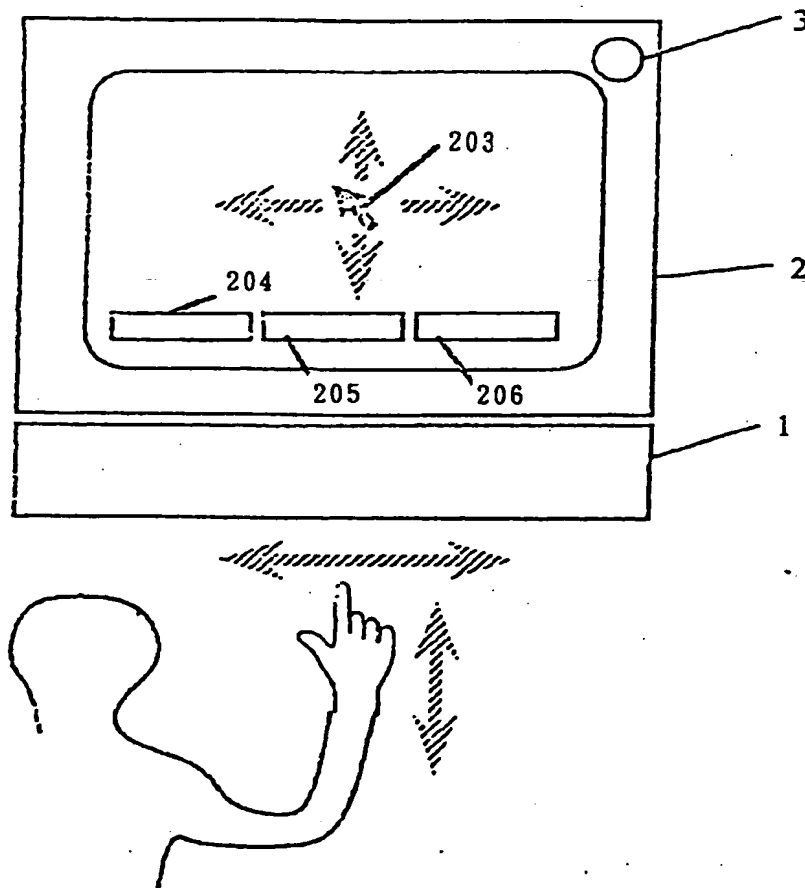


Fig. 8



- 1 Host computer
- 2 Display
- 3 CCD camera
- 203 Icon
- 204. 205. 206 Virtual switches

Fig. 9

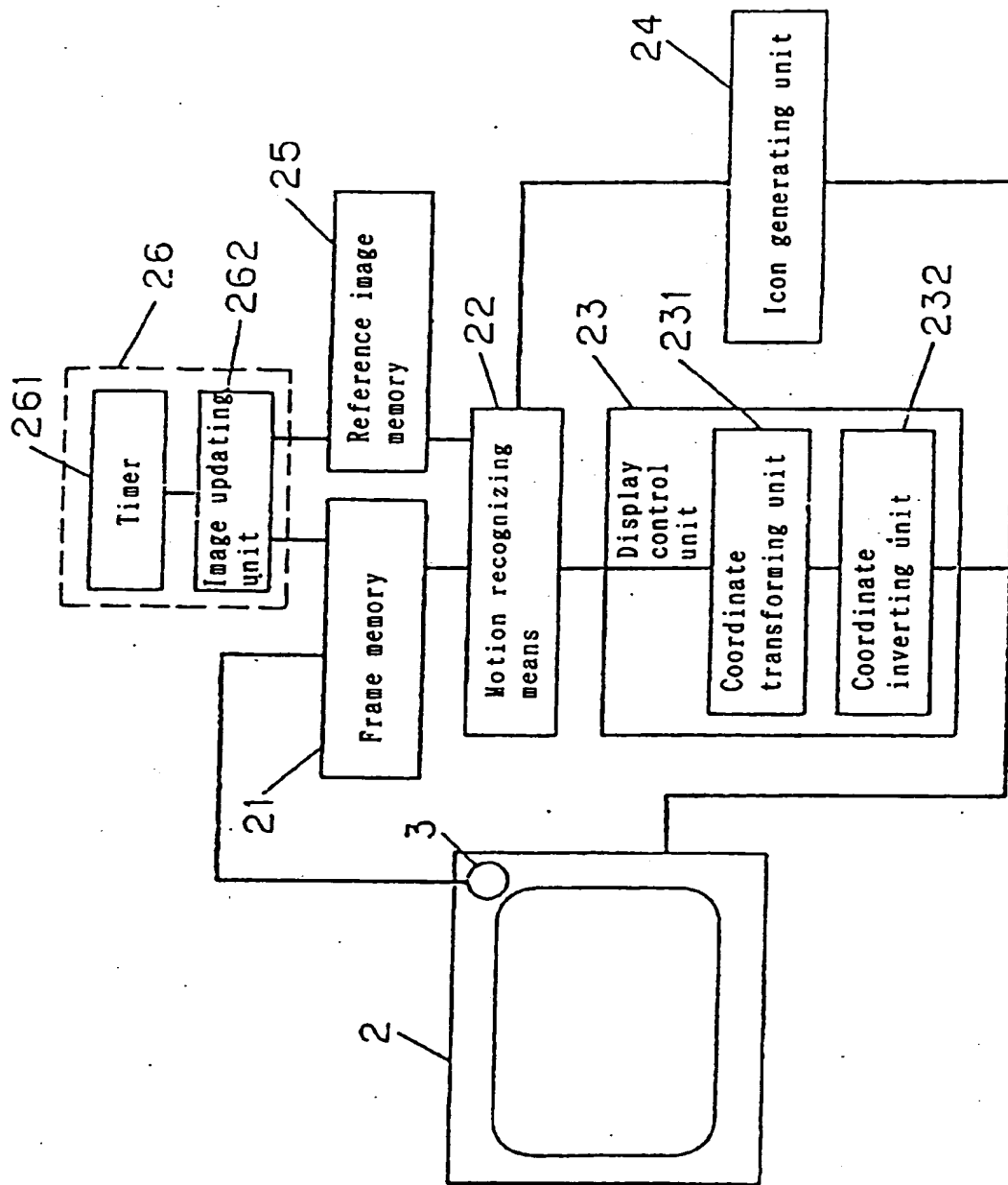


Fig. 10

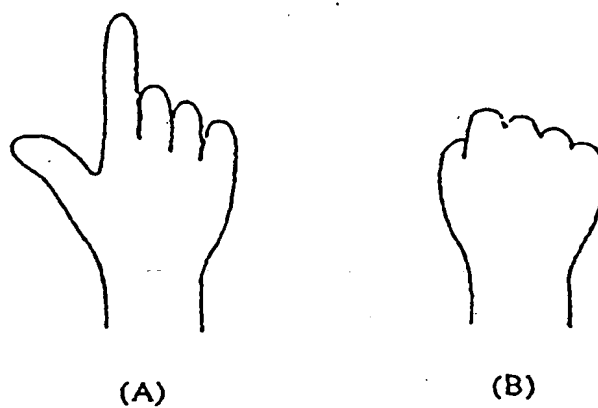


Fig. 11

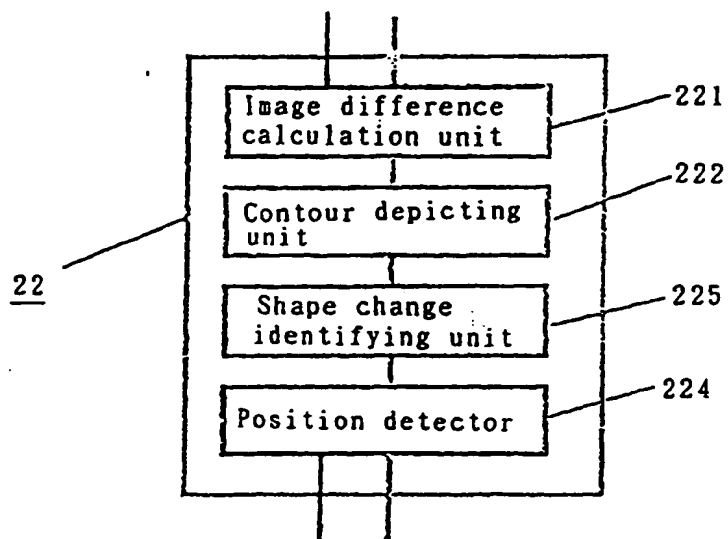


Fig. 12

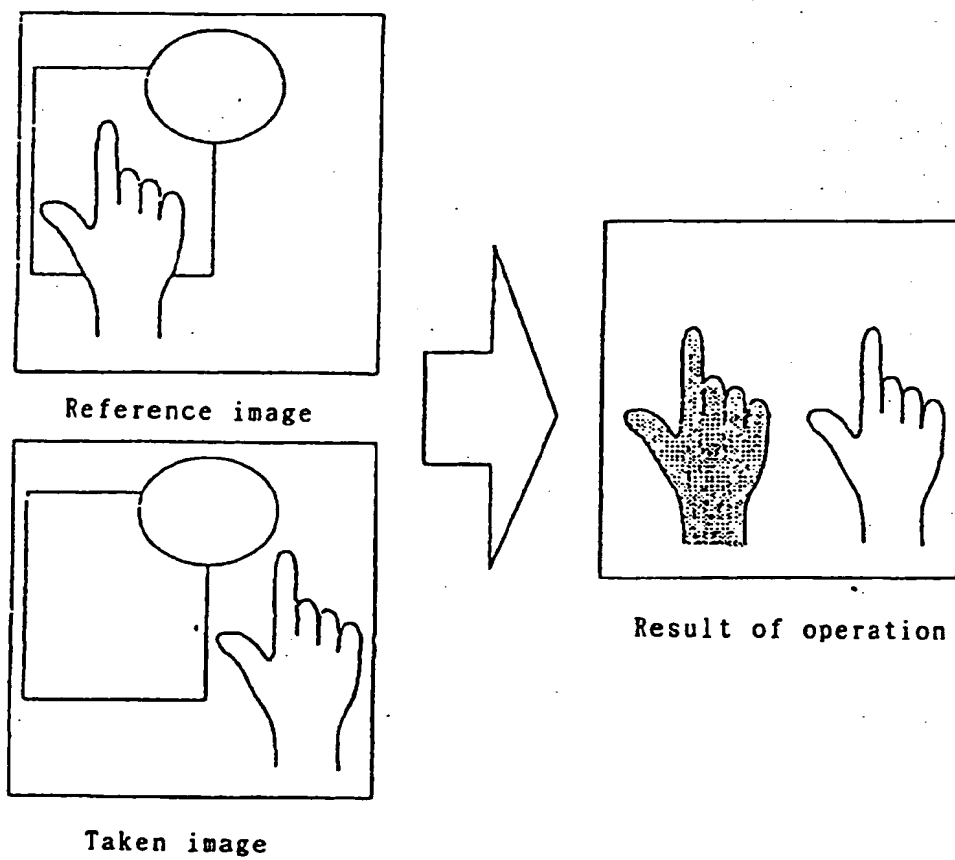
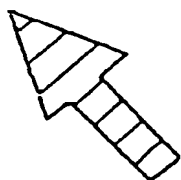
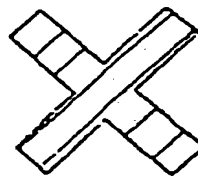


Fig. 13

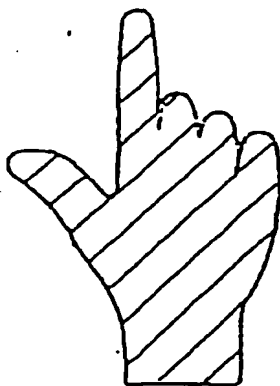
(A)



(B)



(C)



(D)

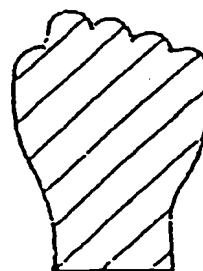


Fig. 14

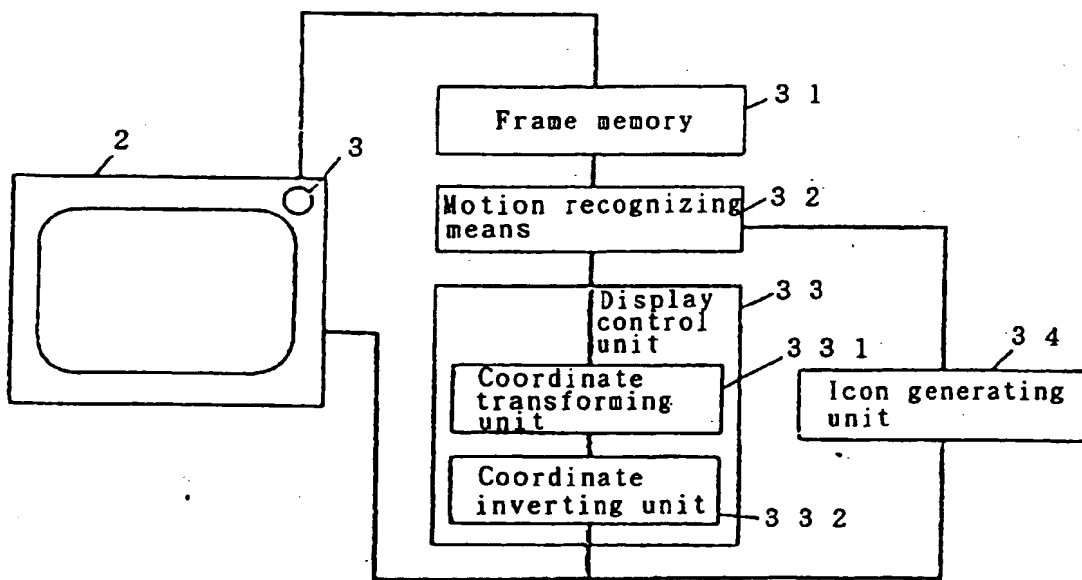


Fig. 15

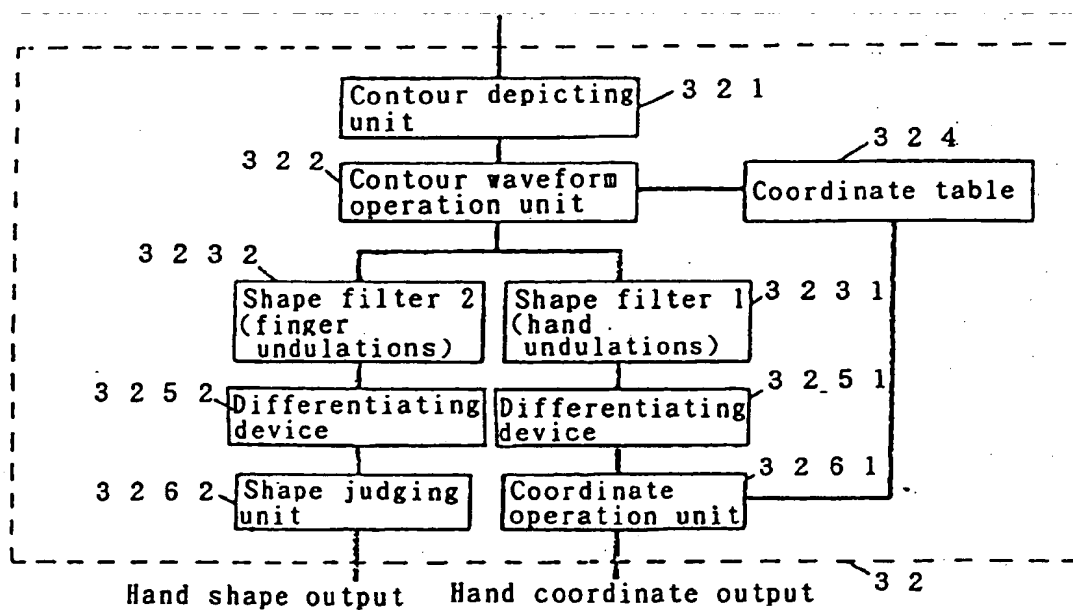
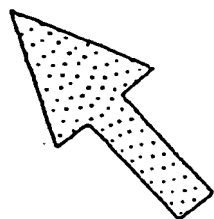
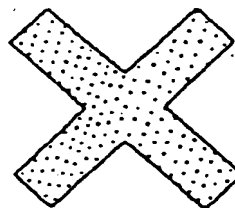


Fig. 16

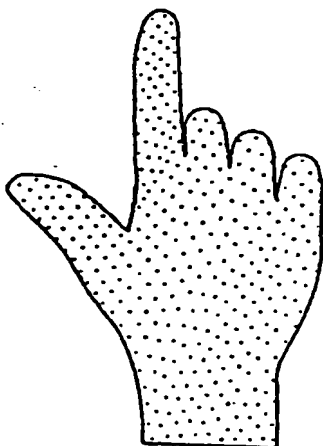
(A)



(B)



(C)



(D)

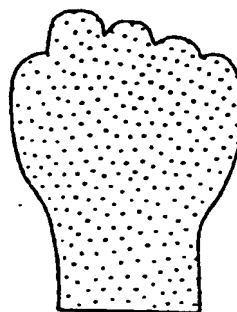


Fig. 17

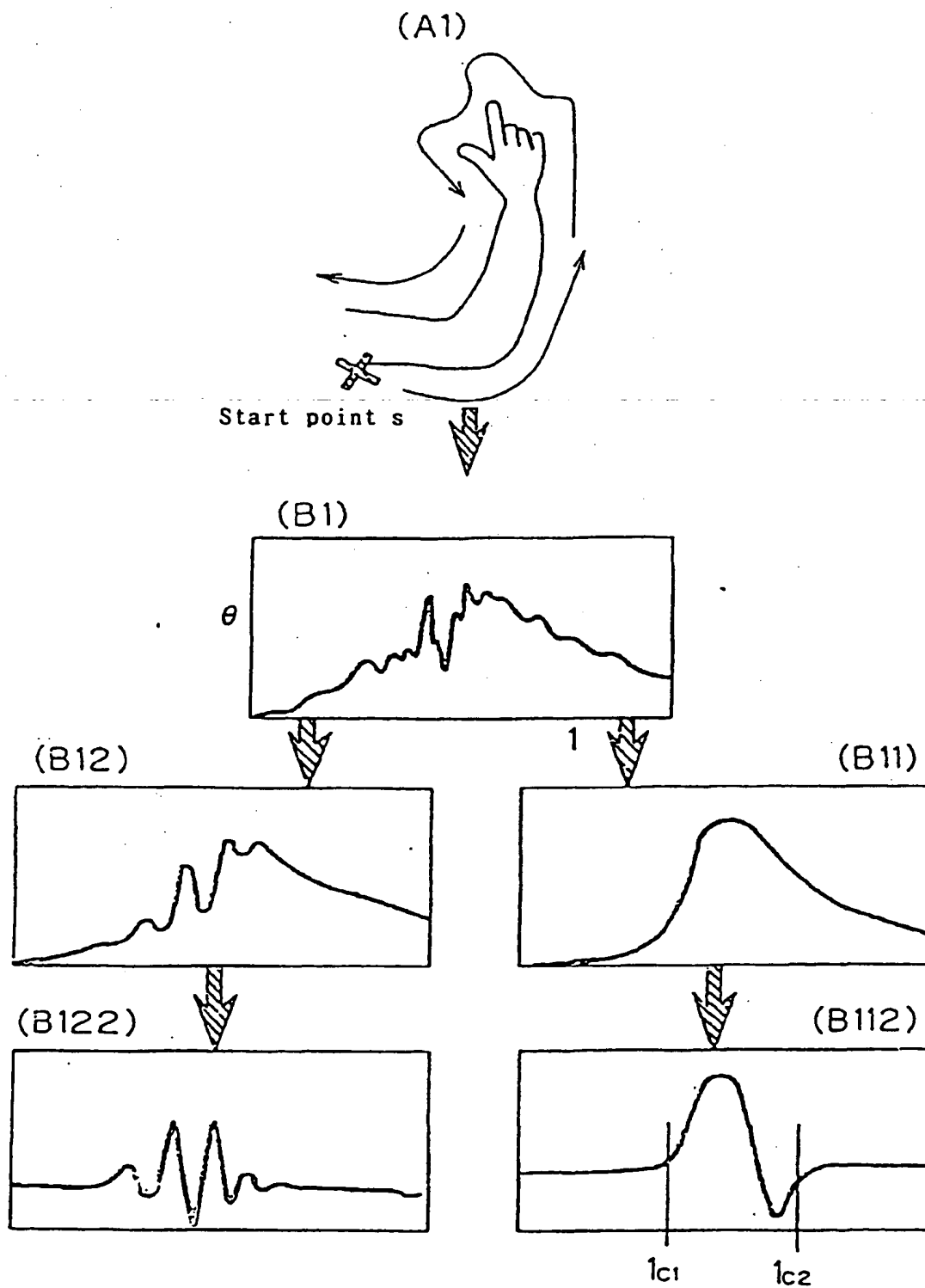


Fig. 18

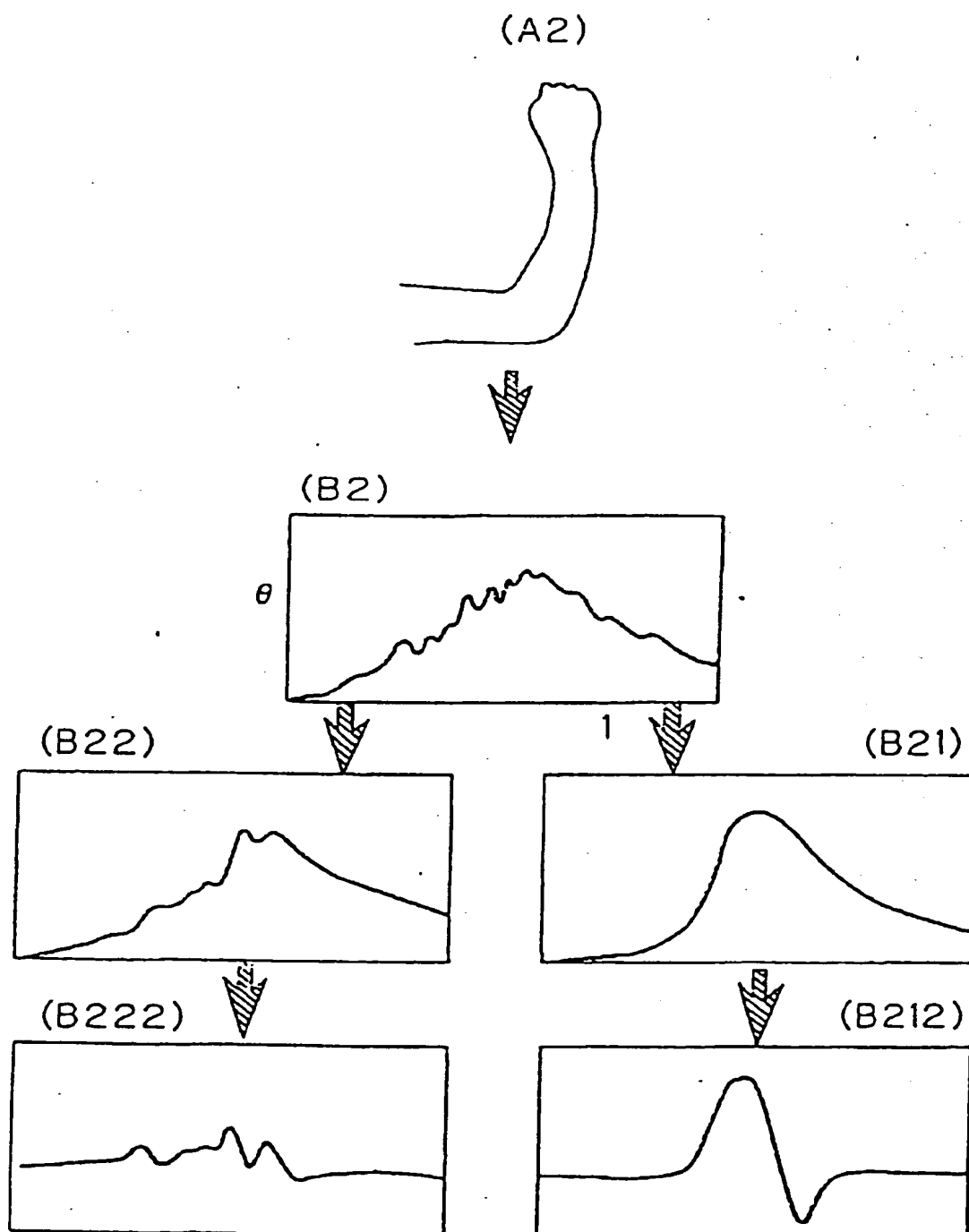


Fig. 19

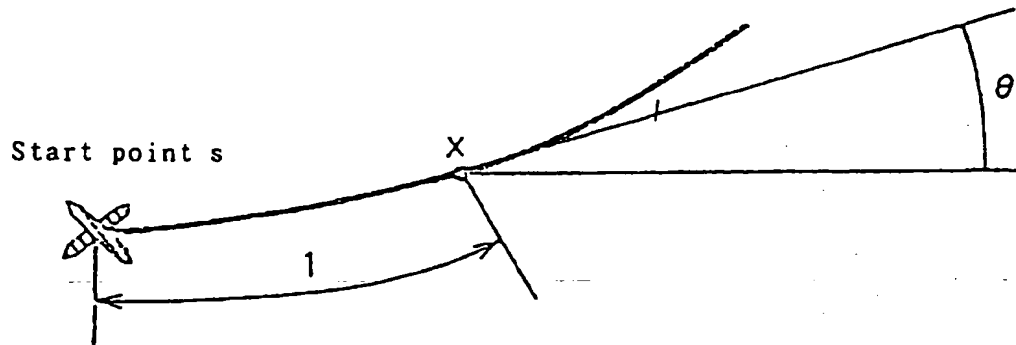


Fig. 20

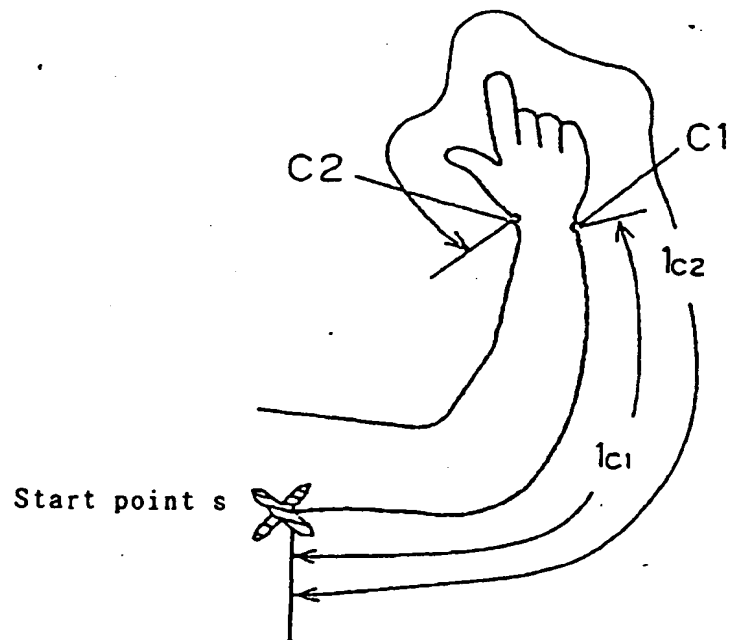


Fig. 21

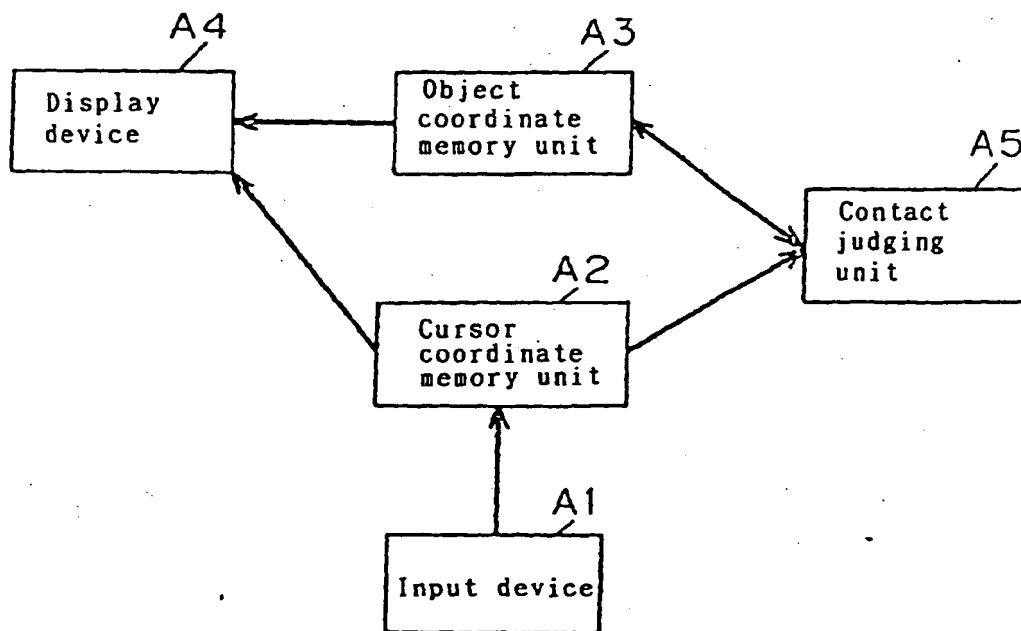
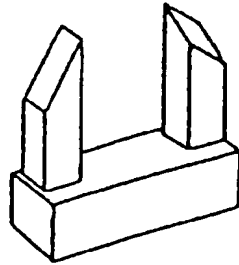
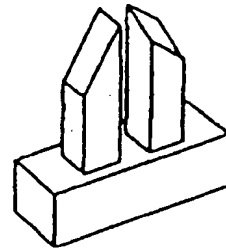


Fig. 22

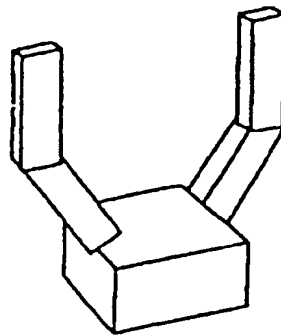
(a)



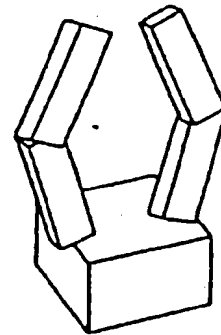
(b)



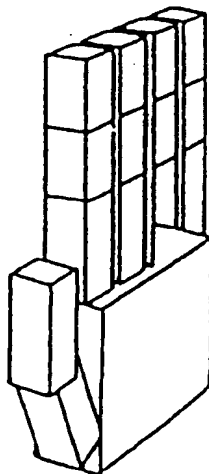
(c)



(d)



(e)



(f)

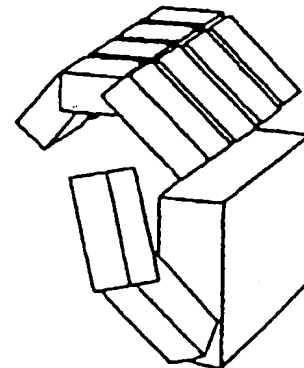
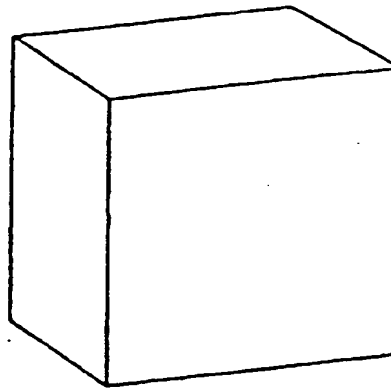


Fig. 23

(a) Cube



(b) Plane

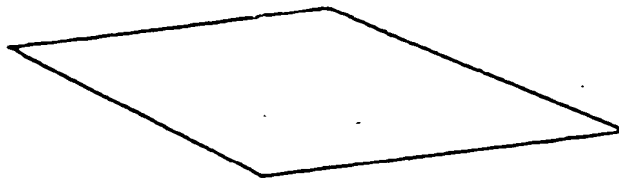
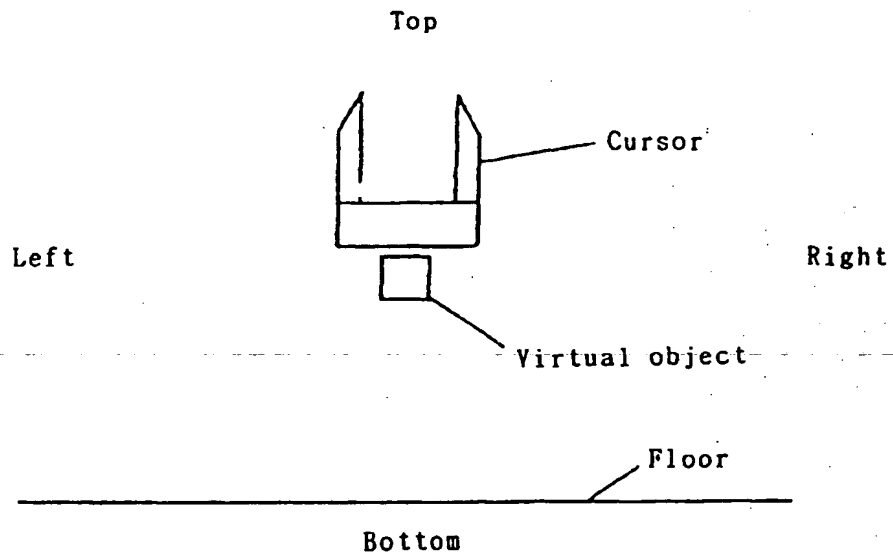


Fig. 24

(a) Example of configuration on virtual space (front view)



(b) Example of configuration on virtual space (side view)

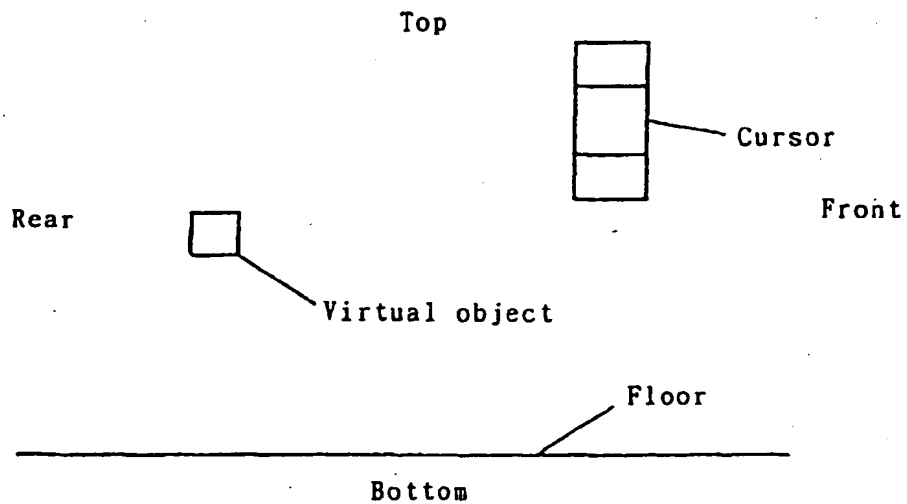


Fig. 25

Display example

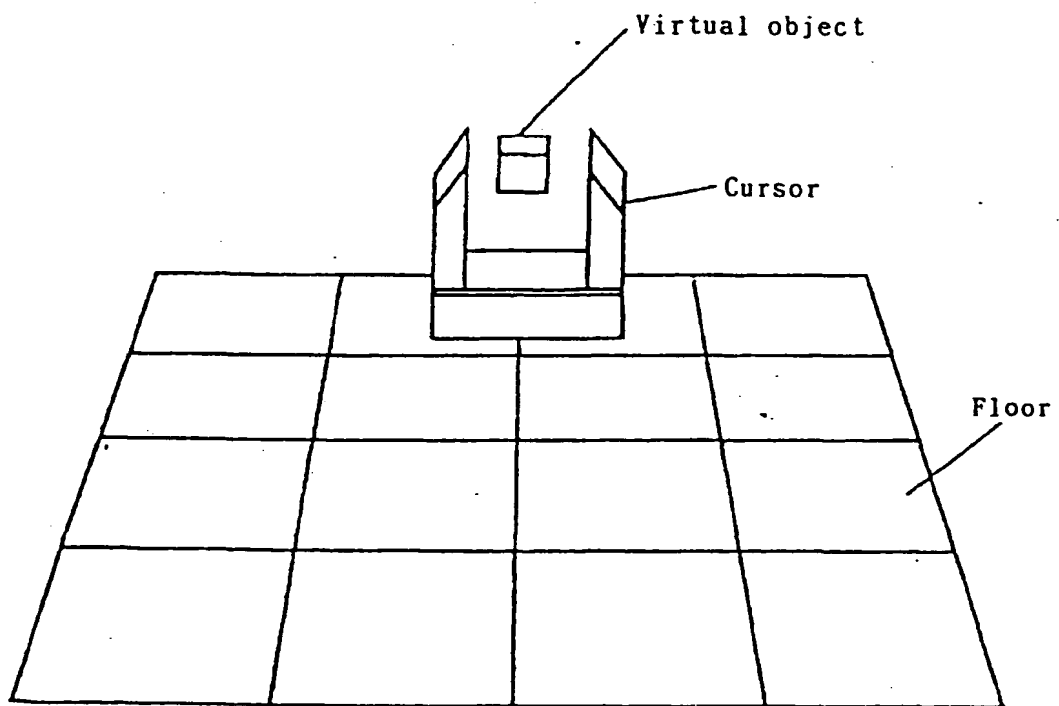


Fig. 26

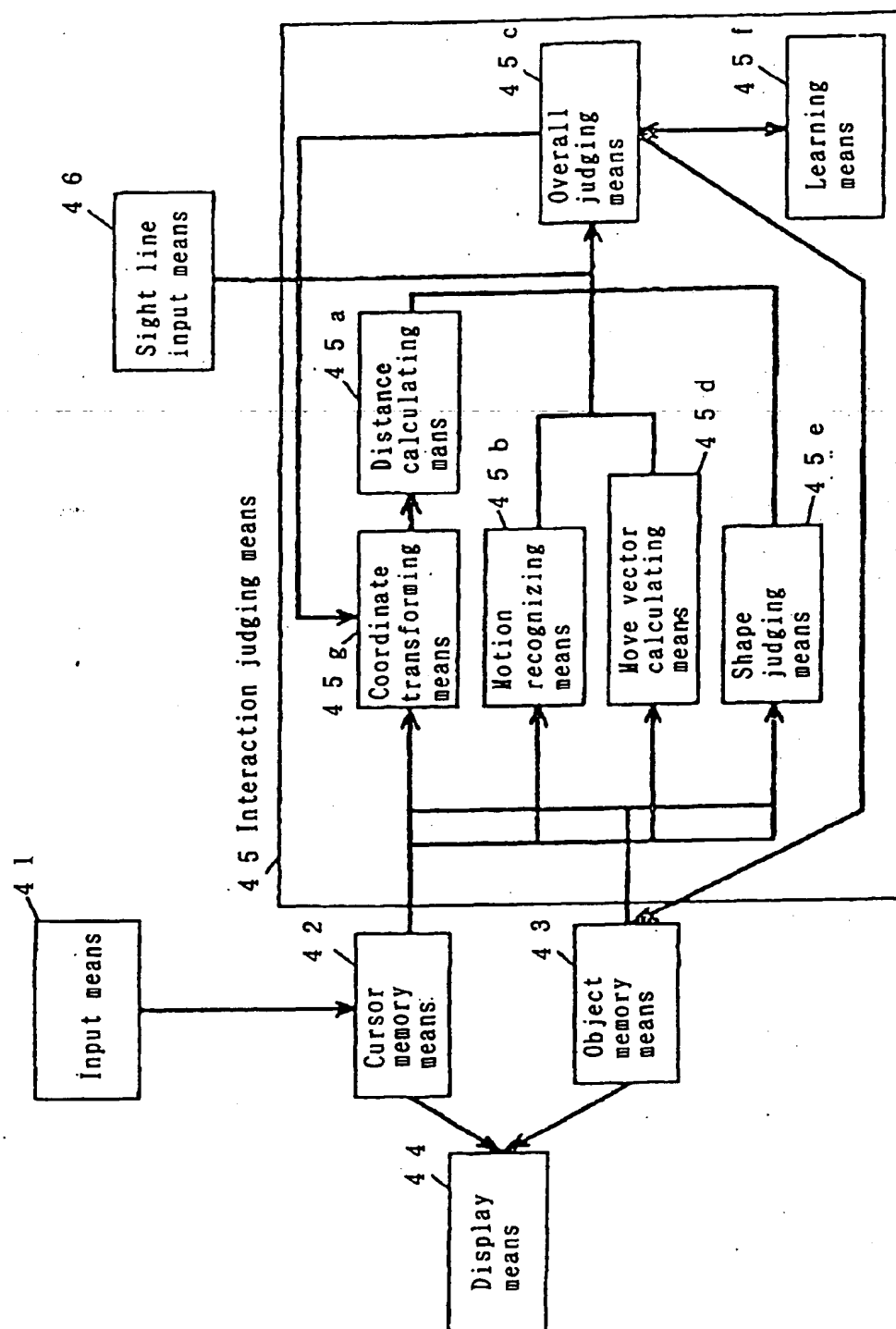
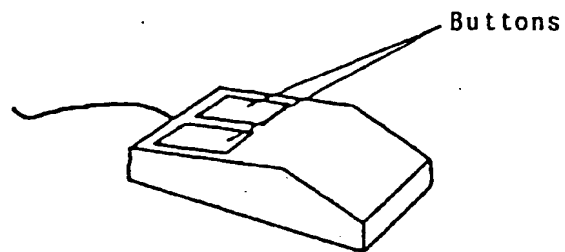
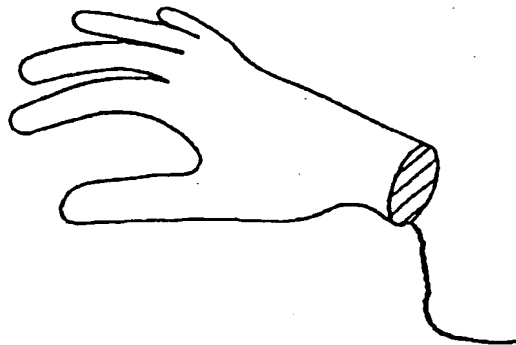


Fig. 27

(a)



(b)



(c)

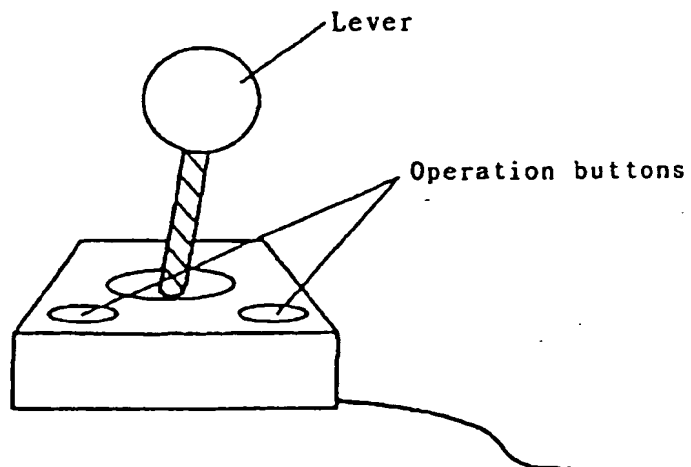
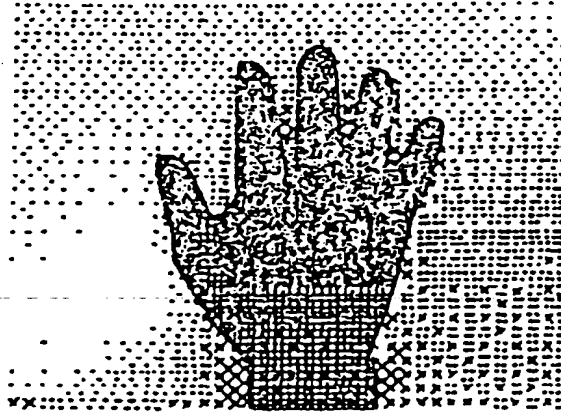


Fig. 28

(a)



(b)

Circumscribing rectangle

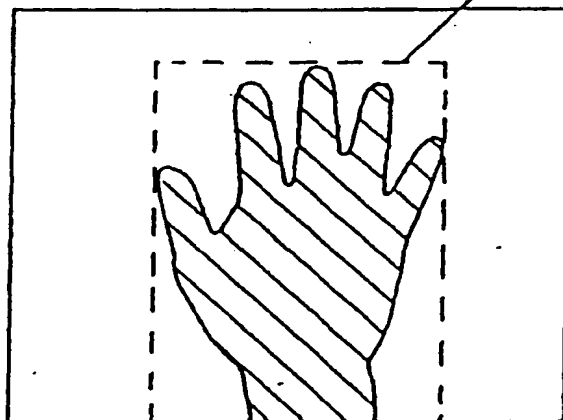


Fig. 29

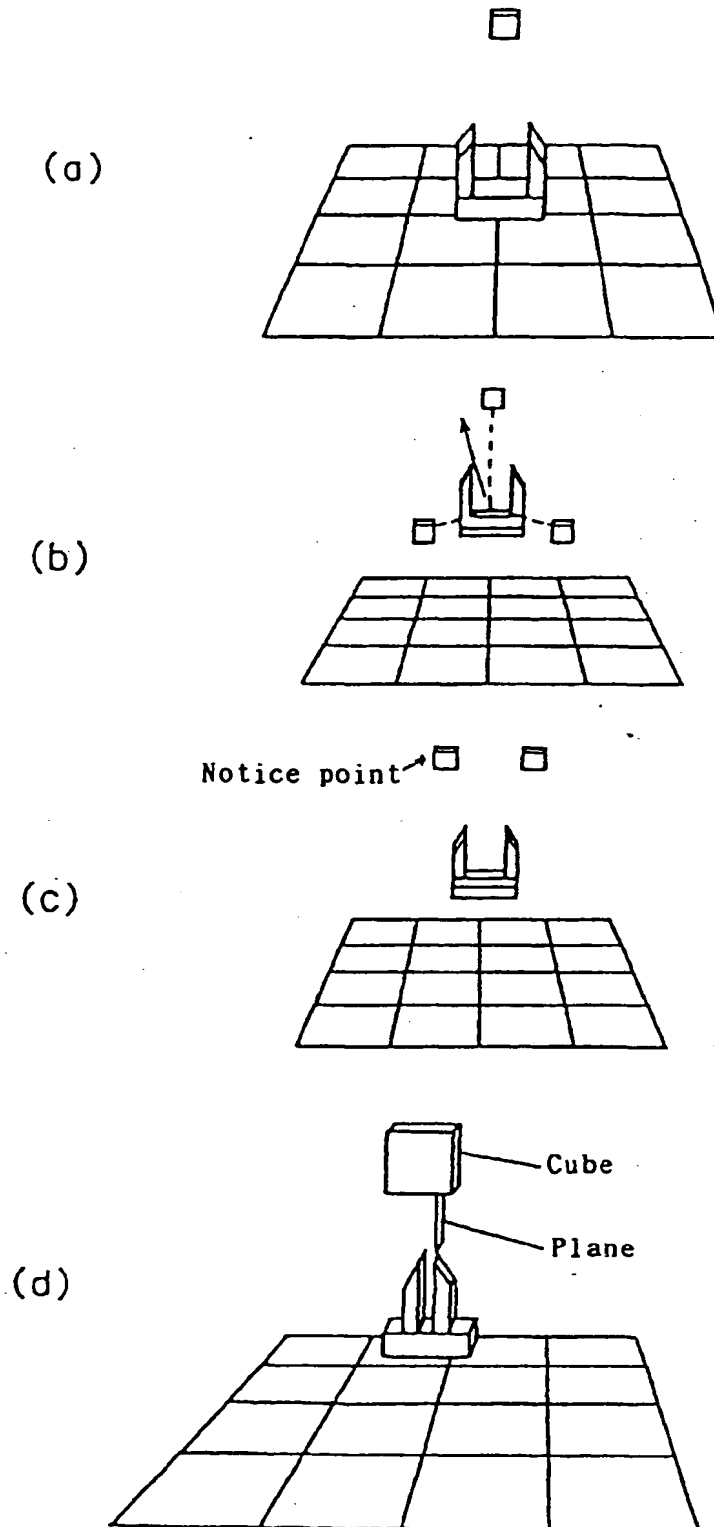
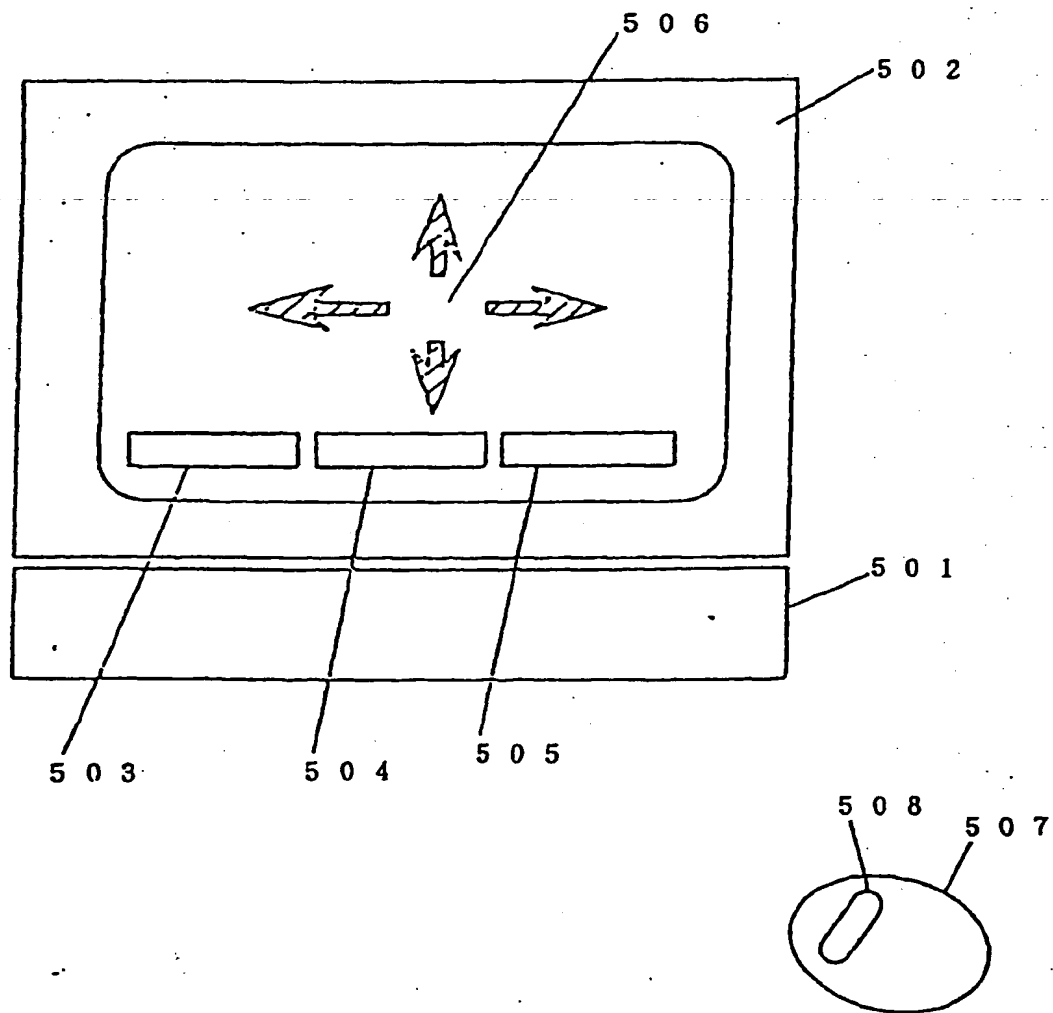


Fig. 30



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP96/01124

A. CLASSIFICATION OF SUBJECT MATTER Int. Cl ⁶ G06F3/033 According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) Int. Cl ⁶ G06F3/033, G06T7/60 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Jitsuyo Shinan Koho 1926 - 1996 Kokai Jitsuyo Shinan Koho 1971 - 1996 Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP, 7-84714, A (Casio Computer Co., Ltd.),	1 - 6
Y	March 31, 1995 (31. 03. 95), Claim; Fig. 4 (Family: none)	7 - 25
X	JP, 6-20055, A (Sony Corp.),	1 - 3
Y	January 28, 1994 (28. 01. 94)	4 - 25
X	JP, 7-57103, A (Toshiba Corp.),	1 - 3
Y	March 3, 1995 (03. 03. 95) (Family: none)	4 - 25
X	JP, 5-66908, A (NEC Corp.),	1-3, 13, 14
Y	March 19, 1993 (19. 03. 93)	4-12, 15-25
X	JP, 6-12177, A (Canon Inc.),	1-3, 13, 14
Y	January 21, 1994 (21. 01. 94)	4-12, 15-25
Y	JP, 2-190776, A (NEC Corp.), July 26, 1990 (26. 07. 90), Fig. 3 (Family: none)	7 - 12
Y	JP, 4-238524, A (Matsushita Electric Ind. Co.,	13-15, 17
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search July 29, 1996 (29. 07. 96)		Date of mailing of the international search report August 13, 1996 (13. 08. 96)
Name and mailing address of the ISA/ Japanese Patent Office Facsimile No.		Authorized officer Telephone No.

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP96/01124

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	Ltd.), August 26, 1992 (26. 08. 92)	
Y	JP, 5-27906, A (Matsushita Electric Ind. Co., Ltd.), February 5, 1993 (05. 02. 93), Fig. 15	13, 14
Y	JP, 4-289910, A (NEC Kobe Software K.K.), October 14, 1992 (14. 10. 92)	20, 21

Form PCT/ISA/210 (continuation of second sheet) (July 1992)



XP 000029910

606K11/18

THE TRACKING CUBE: A THREE-DIMENSIONAL INPUT DEVICE

pg 1-95

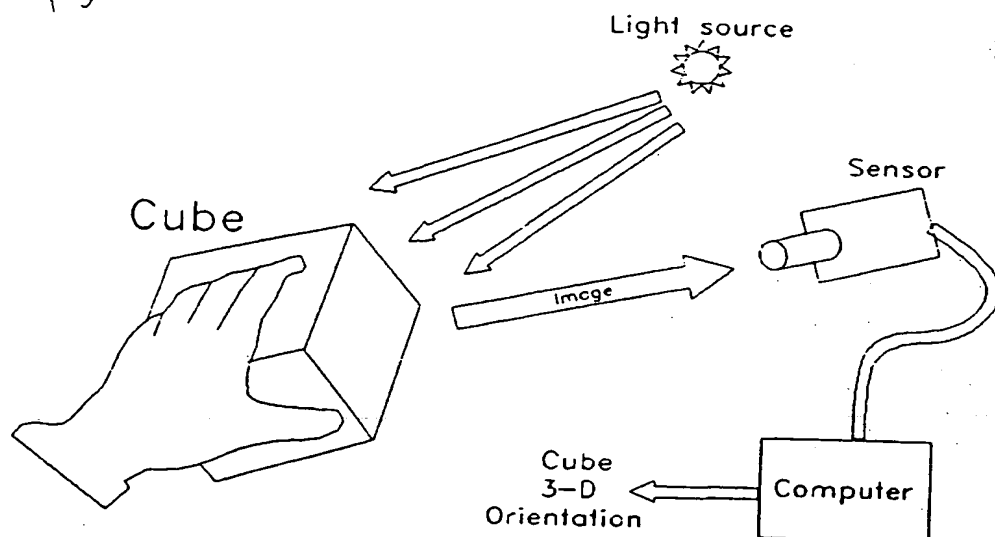


FIG. 1

This article describes a hand-held input device for a graphics station which is connected to a computer. This device allows the entering of 3-D information at the graphics station; it has no physical connection to the computer, but is scanned by a sensing device which calculates its 3-D orientation and returns the result as input to the graphics station.

The proposed input device for 3-D orientation is a tracking cube. Although being an independent device from the screen, it is intended to give the user a strong sensation of being in direct physical control of the displayed graphical 3-D object. Fig. 1 is a schematic illustration of a scenario for using the tracking cube. This device gives an intuitive feedback to the user and the illusion of actually holding in his hand the manipulated 3-D graphical object. In other words, the cube is a real physical object held in space, which represents an imaginary graphical object on the screen. The system keeps updating the 3-D position of the imaginary object in accordance with the physical position of the tracking cube as manipulated by the user.

The tracking cube is made of two parts: a hand-held device which represents the manipulated graphical object, and a sensor which scans the hand-held device and computes continuously its 3-D geometric orientation. The 3-D orientation is computed relative to the coordinate

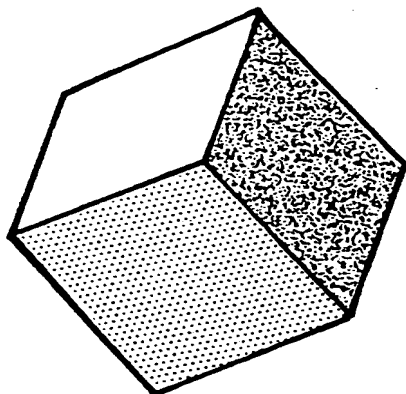


FIG. 2A

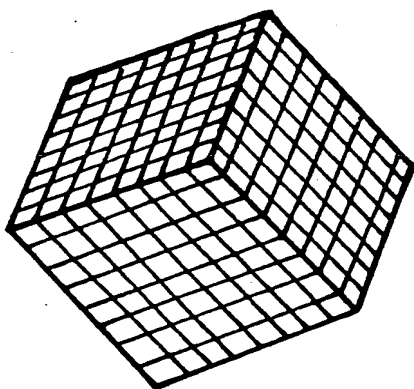


FIG. 2B

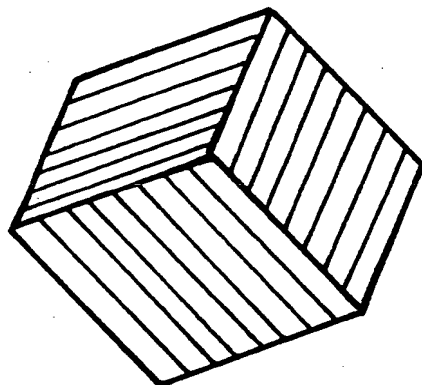


FIG. 2C

system of the sensor. The hand-held device is a cube whose faces are painted specifically to provide cues for surface-orientation which - combined together - can serve to find determinatively the 3-D orientation of the cube. This means that the orientation of each surface cannot be found determinatively on its own. The sensor is a TV scanning camera which records a picture of the cube while held in space, which is then subject to digital processing. This is done repeatedly as fast as possible, preferably at the scanning speed of the camera (50-60 times a second), but resolution of the orientation problem 10-20 times a second will do as well. For the sake of robustness and ease of use, the sensor should not be sensitive to the fact that the cube is held by hand, which partially obscures the orientation cues.

Surface orientation is defined as the unit normal vector to the surface plane. Therefore, the orientation of a surface in 3-D is a 2-D space on its own. If the normal vector is $N = [u \ v \ w]$, then it can be represented by the 2-D space $u-v$.

Object orientation is defined by the relative orientation of its object-centered coordinate system, to the world coordinate system. It is more than sufficient to define two of the objects' axes, say, Z and X. A more intuitive representation is by means of the three rotational transformations to be applied to the object in order to get to its orientation: sequence of rotations around its X, Y and Z axes.

Surface orientation can be determined from a shading model based on the intensity of light reflections from the different faces or from texture models using specific textures on the different faces. Both models are not determinative for a single face, but when there are more than one surface of a single solid object, they can be combined to solve a determinative problem.

Shading Models

For the shading model approach, a cube with diffusely reflecting (smooth) surfaces is used. Fig. 2A indicates a light reflection image of a smooth-surface cube. With shading models, measuring the light reflectance intensity constrains the $u-v$ surface parameters to a 1-D parametric subspace of the $U-V$ plane. When a single light source illuminates the tracking cube along one of its major diagonals, all surfaces will reflect equivalently, and the problem still remains unsolved. Therefore, more than one light source should be used. The proposal here is to use three different light sources. When the light sources can be distinguished, there is more than enough information to solve the cube orientation problem.

Formally, the system takes three images $I_i (i=1,2,3)$ which correspond to the three light sources. Each light source shines its light in the direction of the unit vector $V_i (i=1,2,3)$, and is reflected from the three visible cube faces in the quantities $C_{ij} (i=1,2,3; j=1,2,3)$. These quantities satisfy the equations $E_{ij} : C_{ij} = V_i \cdot N_j$, where N_j is a unit normal vector to the j -th face of the cube, assuming a Lambertian surface model which reflects light equivalently in all directions in an intensity which depends only on the angle between surface normal and lighting direction. The V_i are known, so one can solve for N_j using two of the equations C_{ij} for any particular j . It is suggested to use, however, all three equations for a least-square approximation of the result. To find the cube's orientation, it is sufficient to know only two of the N_j , yet again, sometimes only two faces are clearly visible, and sometimes only one.

To increase the system's robustness, one can do the following:

- The linear system of equations is solved very rapidly, so that the previous state of the cube is not going to change dramatically. Therefore, the linear system of equations should result in a solution not far from the previous state, which is serving as a starting point.

- In some cases, for example, when only one face is visible, not enough information is present in light reflectance alone, and some image processing is needed to find the outline of a square face. This requirement may be expensive, for which the scheme described below under the heading "Texture Models" comes in very handy.

To distinguish the reflections from the different light sources (i.e., to identify correctly the three images), one can use either color coding or time multiplexing.

- a) Color coding means that each light source shines a different color which is easy to identify. Such colors will be those which are farthest away from each other in the Munsel color space. By an obvious rationale, it is better to select colors which correspond mostly with the RGB (Red Green Blue) color sensitivity of the imaging system (color TV camera). The advantage of color coding is that the imaging process is continuous and all colored images are taken simultaneously of a fixed-position cube.

- b) The time multiplexing approach is to strobe the cube from more than one direction in a synchronized sequence with the imaging system, thus collecting more than one image of the cube at each step. The imaging system can still employ a monochrome (and less expensive) device, although the images are not taken at the exact same time, and there might be some variations in the cube position. With the regular scanning frequency of 50 Hz, each image takes about 20 ms, so that three images will span more than 20th of a second. The rate of resolving the orientation problem can still be as fast as the scanning frequency (50-60 Hz) when each time the three most recent exposures (which are for three different directions) are used.

Texture Models

With texture models, texture elements can provide more than just a single constraint of the u-v parameters. By properly choosing the texture pattern, orientation of the texture elements on the surface can be determined and used. Fig. 2B shows a cube with a grid pattern on all faces. With the proper determination of these texture elements, enough information may be gathered to solve the orientation problem. Fig. 2C is another proposal for texture coding the cube faces so that each face has a simple and unique texture distinguishable easily from all its neighbors' faces. This texture code is superior since it allows an easy analysis of the image for calculating the cube orientation parameters. For these texture models, the cube is only illuminated by a single light source.

A scheme for computing surface orientation from the striped texture (Fig. 2C) is to do an FFT on the image, and isolating peaks in the resulting image. These peaks correspond to major spatial frequencies in the original image. Since the stripes are a dominant feature of the image, all other spatial frequencies (caused by a holding hand, for instance) will be much weaker. Therefore, the three major directions of the groups of stripes will be easily isolated. Based on this information, the cube orientation can be computed. Even when only one cube face is visible, this scheme will not require any extra expensive image processing (like line detection with the shading models) to find cube orientation.

Other Stylus Shapes

The stylus held in hand can have other shapes besides a cube:

1. Polyhedra: Any polyhedron which has planar faces can be used for all of the color, reflection, and texture coding approaches described above. The geometrical relation between planar faces will govern the equations to be used. Yet, the cube seems to be the best and most perfect candidate of the polyhedra family, due to its nicely employed symmetry. For shading coding, however, by using higher order polyhedra, one can eliminate the case of only one or two visible faces.
2. Spherical shapes: A spherical shape can be obtained by "blowing" the cubical shape. A spherical shape will lose the option of using light reflectance-based techniques, as well as color-coding, too. Yet, texture coding can still be applied.

